NASA TECHNICAL NOTE

NASA TN D-4272

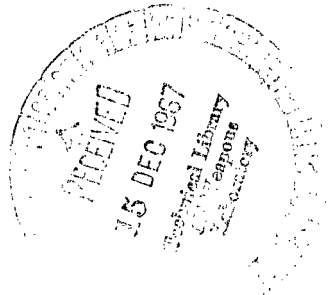
MULTIPLE DECISION PROCEDURES FOR ANOVA OF TWO-LEVEL FACTORIAL FIXED-EFFECTS REPLICATION-FREE EXPERIMENTS

by

Arthur G. Holms
Lewis Research Center

and

J. N. Berrettoni
Western Reserve University



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • DECEMBER 1967

NASA TN D-4272

MULTIPLE DECISION PROCEDURES FOR ANOVA OF TWO-LEVEL FACTORIAL FIXED-EFFECTS REPLICATION-FREE EXPERIMENTS

By Arthur G. Holms

Lewis Research Center
Cleveland, Ohio

and J. N. Berrettoni

Western Reserve University
Cleveland, Ohio

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

MULTIPLE DECISION PROCEDURES FOR ANOVA OF TWO-LEVEL FACTORIAL FIXED-EFFECTS REPLICATION-FREE EXPERIMENTS*

by Arthur G. Holms and J N. Berrettoni†

Lewis Research Center

SUMMARY

For expensive areas of experimentation, such as alloy development, pressure vessel burst testing, and high-temperature protective coatings, the appropriate experiments consist of two-level fixed-effects factorial designs without replication. No adequate procedures have been available for the statistical analysis of such experiments.

A procedure called "chain pooling" is introduced for testing the significance of terms of a model equation as fitted to the observations from a fractional factorial experiment. The procedure starts with a small group containing only the smallest of the ordered squared coefficients of the model equation in the denominator of a test statistic. Stepwise testing, in the increasing order of succeeding squared coefficients, pools insignificant squares into the denominator of the statistic, which is used for continued testing.

Monte Carlo computations were performed to determine the decision error probabilities for many different variations of chain pooling and to compare the relative advantages of the variations for fractional factorial experiments with 2^4, 2^5, and 2^6 treatment combinations. These computations were performed with values of the significant coefficients distributed in such a manner as to contribute to high probabilities of decision errors, so that the recommended procedures are good against the worst possible conditions.

For any real experiment, the actual decision error probabilities depend on the magnitudes of the coefficients. A method is given for estimating upper limits of weighted average decision error probabilities after the coefficients have been estimated. The

*This report is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosphy, at Western Reserve University, Cleveland, Ohio, June 1966.

†Chairman, Department of Statistics, Western Reserve University, Cleveland, Ohio.

procedures are illustrated by an example from high-temperature-alloy development.

INTRODUCTION

The factorial experiment is useful for observing the response of a continuous dependent variable to changes in the independent variables. The independent variables may be continuous, discrete, or qualitative. The factorial experiment is preferred to other designs when certain combinations of levels of the independent variables can affect the response (interact).

If independent variables x_1, x_2, x_3, . . . are to be investigated at numbers of levels a, b, c, . . ., and if the error is to be measured with an r-fold replication, the full factorial experiment includes $r \cdot a \cdot b \cdot c \ldots$ observations. A replicated full factorial experiment can be too expensive in such fields as alloy development, destructive tests of structures (liquid rocket fuel tank bursting), or where many variables are involved (high-temperature protective coatings).

Identifying those factors that affect the response directly or through interactions is done efficiently by performing the experiments at only two levels of each factor. A full factorial experiment on g factors then requires 2^g observations and provides estimates of the direct effects and all interactions.

The interactions involving the larger numbers of factors are often anticipated to be negligible, and the experiments are then performed as fractional replicates. A fraction $(1/2)^h$ of the full factorial experiment is performed. The number of observations is

$$2^{\ell} = 2^{g-h}$$

and $\ell = g-h$.

The 2^{g-h} experiment is preferred when each experimental unit is costly but where any existing interactions should be discovered. The basic design could involve $r2^{g-h}$ observations, where an r-fold replication is used to estimate the error variance. The economy achieved by not replicating (by setting $r = 1$) carries the penalty that there is no obvious, or prior, valid mean square for estimating the error variance, and an estimate of error variance is needed in selecting those effects that will be judged significant.

If replication is lacking, a customary practice, according to Davies (ref. 1, p. 286), consists of pooling some arbitrary number of the highest order interaction mean squares into an estimate of error variance. However, if this practice is followed, any unknown block effects could inflate some of the pooled interactions and thereby give too large an

estimate of error variance. Too large an error estimate reduces the sensitivity of subsequent tests to detect real effects among the estimates of the main effects and lower order interactions.

The preservation of sensitivity, when pooling mean squares into the estimate of error variance, has been an object of the procedure of Daniel (ref. 2) and of Wilk, Gnanadesikan, and Freeny (ref. 3). Daniel uses the absolute values of the effect estimates as order statistics. These values are plotted on probability paper and the result is called a half-normal plot. Such a display, combined with a background of experience, might provide a method by which a skillful user could pass judgment on the results of an experiment. Daniel concluded that the half-normal plot can be used to make judgments about the reality of the largest effects observed only if a small proportion of the effect estimates represent real effects.

Birnbaum (ref. 4) investigated procedures related to half-normal plotting. His results on significance are limited to the single largest order statistic. He concluded, however, that such procedures are optimal with respect to the two largest order statistics.

For the 2^{ℓ} experiment, and aside from the grand mean, there are $2^{\ell} - 1$ mean squares requiring decisions as to significance. The procedure of Wilk, Gnanadesikan, and Freeny (ref. 3) requires that some subjective or prior knowledge be used to decide that η of the $2^{\ell} - 1$ mean squares do not contain real effects and that, therefore, $\rho = 2^{\ell} - \eta - 1$ mean squares do contain real effects. As shown in reference 3, the procedure is not robust against errors in guessing the value of η, and η must be guessed because it is an unknown in the problem.

Daniel and Birnbaum have limited their results to experiments where only a small proportion of the effects are thought to be significant. On the other hand, situations can exist where the experimenter might design a two-level factorial experiment so that a large proportion of the effects will be significant. A particular example occurs in the development of superalloys. These alloys typically contain 5 to 15 elements. One procedur for finding an optimum composition is to use Box-Wilson techniques (ref. 1, p. 495). The first phase fits a first-degree response model to data gathered from a factorial experiment. The costs of experimenting and the need to investigate many elements imply that experiments should be fractionally replicated. Efficiency of the fractional design requires that most of the degrees of freedom (in fitting the linear model) be assigned to the direct effects with only a few contrasts assigned to the interactions. The Box-Wilson techniques also imply that the experimenter will achieve conditions where the first-degree model is no longer valid. He needs a method for deciding to abandon the first-degree model. One such method regards those interactions that are evaluated as a sample of higher degree effects. If they give evidence of higher degree, the experimenter performs the more extensive experiments required by the second-degree model. (Se-

quences of blocked fractional designs, especially appropriate to proceeding in steps from the first-degree model to the second-degree model, were presented in ref. 5.)

The observed effects from the two-level experiment are thus used for three purposes:

(1) To estimate error variance

(2) To evaluate main effects and test their significance

(3) To evaluate interaction effects and test their significance

In alloy development, the metallurgist often has enough prior knowledge to set composition levels so that most of the main effects will be significant. Consequently, the testing of the sample of interactions must be based on an error variance estimate that comes from a small (but not predetermined) number of nonsignificant effects. Decision procedures are needed that will use a small conditional number of effects to estimate error variance. Such procedures were not provided by Daniel or Birnbaum.

The method proposed herein called chain pooling, tests a major proportion of the mean squares in the order of increasing magnitude. Certain procedures will be proposed as being reasonable. Their properties will not be investigated analytically; however, appropriate risk functions will be defined and several variations of the suggested procedures will be evaluated by Monte Carlo methods in terms of the risk functions. The investigation will be limited to experiments that are of 2^4, 2^5, and 2^6 fractional factorial design.

SYMBOLS

C_n	Cochran's statistic for largest of j mean squares
$E(\ldots)$	expectation of $(\ldots)$
e	single observation random error
g	number of factors in two-level experiment
h	fractional replicate contains $(1/2)^h$ observations of full factorial experiment
i	subscript denoting order of computing mean squares according to Yates' algorithm; $i = 0, 1, 2, \ldots, 2^{\ell} - 1$
j	subscript denoting j^{th} smallest mean square (exclusive of grand mean) $j = 1, 2, \ldots, 2^{\ell} - 1$
k	subscript
$L(\lambda_i, d)$	loss for decision d under parameter λ_i

ℓ	experiment contains 2^{ℓ} treatment combinations and produces that many observations
m	number of mean squares pooled before testing begins
N	number of Monte Carlo generated experiments
N(. . .)	normal distribution with parameters (. . .)
n	sample size
P_{1i}	observed probability of type 1 error when testing i^{th} mean square
P_{2i}	observed probability of type 2 error when testing i^{th} mean square
R_1	risk associated with type 1 errors
R_2	risk associated with type 2 errors
r	number of replications
U_j	test statistic defined by eq. (8)
V(. . .)	variance of (. . .)
x	levels of independent variables
Y	response
Z	mean square
α	test size
α_f	nominal size of final significance test
α_p	nominal size of preliminary pooling test
β	type 2 error probability
δ_i	parameter determining relative magnitudes of real effects in any one experiment
ς_j	expectation of j^{th} order statistic of a $\chi^2_{(1)}$ variable
η	number of mean squares having noncentrality parameter of zero
$\hat{\eta}$	estimate of η
θ	scale parameter
$\hat{\lambda}$	estimate of λ
λ	average noncentrality parameter
μ_i	coefficients of eq. (1) that are estimated in Yates' order from Yates' contrasts
$\hat{\mu}_i$	estimate of μ_i

ρ	number of real effects
$\hat{\rho}$	estimate of ρ
$\sum(\ldots)$	summation of $(\ldots)$
σ	standard deviation
$\hat{\sigma}$	estimate of σ
$\chi^2(\ldots)$	chi-square distribution with $(\ldots)$ degrees of freedom
ψ	detection efficiency defined by eq. (15)

CHAIN POOLING

Analysis of Variance Model

Consider a 2^{ℓ} experiment with $\ell = 4$. The factors can be qualitative or quantitative and are named x_1, x_2, x_3, and x_4. Their levels are represented by +1 for the upper level and by -1 for the lower level. The model for the response is then written as

$$Y = \mu_0 + \mu_1 x_1 + \mu_2 x_2 + \mu_3 x_1 x_2 + \mu_4 x_3 + \mu_5 x_1 x_3 + \mu_6 x_2 x_3 + \mu_7 x_1 x_2 x_3$$

$$+ \mu_8 x_4 + \mu_9 x_1 x_4 + \mu_{10} x_2 x_4 + \mu_{11} x_1 x_2 x_4 + \mu_{12} x_3 x_4 + \mu_{13} x_1 x_3 x_4$$

$$+ \mu_{14} x_2 x_3 x_4 + \mu_{15} x_1 x_2 x_3 x_4 + e \qquad (1)$$

where

$$E(e) = 0; \quad V(e) = \sigma^2$$

and e is independently $N(0, \sigma^2)$.

The observations from the 2^{ℓ} treatments are used to compute mean squares conveniently by Yates' method (ref. 1, p. 263). Assume that the mean squares Z_i have been computed in Yates' order and in this order are labeled $Z_0, Z_1, \ldots, Z_n$, where $n = 2^{\ell} - 1$. The coefficients of equation (1) are in Yates' order. The expectations of the Z_i are

$$E(Z_i) = \sigma^2 + 2^{\ell}\mu_i^2 \qquad (2)$$

$$i = 0, 1, 2, \ldots, 2^{\ell} - 1$$

The quantities Z_i/σ^2 are either central chi-square or, more generally, noncentral chi-square variables that have 1 degree of freedom (ref. 6, p. 227). Let λ_i be the noncentrality parameter. Then

$$\lambda_i = \frac{2^{\ell}\mu_i^2}{\sigma^2} \tag{3}$$

Risk Functions

Assume that n single degree-of-freedom mean squares are drawn from n populations. An unknown number ρ of the populations have real effects ($\lambda_i > 0$) and the balance are null populations ($\lambda_i = 0$). A number $\hat{\rho}$ of the populations are to be selected with the hope that they will be the populations with $\lambda_i > 0$).

Errors of selection are assumed to produce losses that depend on the parameters λ_i and on the decision d as given by loss functions $L(\lambda_i, d)$. The loss for any correct decision is defined as zero. If the i^{th} population is correctly decided to be null, the type 1 loss is

$$L_1(\lambda_i, d) = L_1(\lambda_i = 0, \lambda_i = 0) = 0$$

If the i^{th} population is correctly decided to be nonnull, the type 2 loss is

$$L_2(\lambda_i, d) = L_2(\lambda_i > 0, \lambda_i > 0) = 0$$

If the i^{th} population is incorrectly selected (null hypothesis incorrectly rejected), a type 1 unit loss is assumed:

$$L_1(\lambda_i, d) = L_1(\lambda_i = 0, \lambda_i > 0) = 1 \tag{4}$$

This definition equates the type 1 risk (expectation of type 1 loss) to the probability of a type 1 error.

A dimensionless parameter θ is defined by

$$\frac{1}{\rho}\sum_{i=1}^{\rho}\left(\frac{\mu_i}{\theta\sigma}\right)^2 = 1$$

Let

$$\delta_i = \frac{\mu_i}{\theta\sigma} \tag{5}$$

Then

$$\frac{1}{\rho}\sum_{i=1}^{\rho} \delta_i^2 = 1 \tag{6}$$

A long tradition exists for saying that losses in estimation are proportional to the square of the error. This tradition is followed by saying that if μ_i is incorrectly decided to be zero, the type 2 loss will be proportional to μ_i^2. The type 2 loss is made proportional to the square of the μ_i (if not selected) by

$$L_2(\lambda_i,\ d) = L_2\left(\lambda_i = \frac{2^{\ell}\mu_i^2}{\sigma^2},\ \lambda_i = 0\right) = \delta_i^2 \tag{7}$$

The type 2 risk (expectation of a type 2 loss) is therefore equal to the probability of a type 2 error multiplied by a weighting factor δ_i^2, where the δ_i^2 is proportional to the square of the μ_i that was not selected, and where the weighting factors have a mean value of 1 for the set of ρ nonnull populations.

Test Statistics

As stated in reference 4, the optimal decision procedure for $\rho \leq 1$ uses a test developed by Cochran (ref. 7). The Z_i in Yates' order (omitting the mean square for the grand mean) are ordered in nondecreasing magnitude as Z_j,

$$Z_1 \leq Z_2 \leq \ldots \leq Z_j \leq \ldots \leq Z_n$$

Cochran's statistic is

$$C_n = \frac{Z_n}{Z_1 + \ldots + Z_n}$$

and the null hypothesis is rejected with test size α if C_n exceeds the upper 100α-percent point of Cochran's distribution, which has been additionally tabulated in reference 8.

Assume that the λ_i with Yates' order are written as λ_j in nondecreasing order of the mean squares. Rejection with Cochran's statistic suggests that $\lambda_n > 0$. If $\lambda_n > 0$ is true, then λ_{n-1} might be tested with the statistic

$$C_{n-1} = \frac{Z_{n-1}}{Z_1 + Z_2 + \ldots + Z_{n-1}}$$

Suppose $\lambda_{n-1} = 0$ is rejected. This suggests that the test of λ_n had very low sensitivity because of the inflated denominator. Obviously, a general multiple decision procedure cannot be developed using Cochran's test in the descending order of the mean squares. In ascending order, let

$$C_2 = \frac{Z_2}{Z_1 + Z_2}$$

and if the test based on this statistic accepts $\lambda_2 = 0$, use the conclusion as an assumption and form

$$C_3 = \frac{Z_3}{Z_1 + Z_2 + Z_3}$$

In other words, assume that the smallest mean square Z_1 has been drawn from a population with $\lambda_1 = 0$. Proceed stepwise with test statistics C_j so long as C_{j-1} indicates $\lambda_{j-1} = 0$. At the first rejection of the null hypothesis (e.g., for C_j), conclude that $\lambda_j > 0$, and because of the ordering, immediately conclude that $\lambda_k > 0$ for all $k \geq j$. Thus, Cochran's test has been generalized to a sequence of dependent tests; furthermore, the number of items that should be in the denominator is always unknown. For these reasons the nominal α of Cochran's test will not be the true size for the multiple procedure.

A trivial transformation of Cochran's statistic provides an alternative statistic. It is

$$U_j = \frac{jZ_j}{Z_1 + \ldots + Z_j} \qquad (8)$$

$$j = 2, \ldots, n$$

The basic process of chain pooling begins with $\lambda_1 = 0$ and $j = 2$. At each stage of the testing, the composite null hypothesis

$$\lambda_1 = , \ldots, = \lambda_{j-1} = \lambda_j = 0; \ 0 < \lambda_{j+1}, \ldots, \lambda_n$$

is tested against the composite alternative hypothesis

$$\lambda_1 = , \ldots, = \lambda_{j-1} = 0; \ 0 < \lambda_j, \ \lambda_{j+1}, \ldots, \lambda_n$$

The null hypothesis is rejected at the first j for which U_j exceeds the tabulated 100α-percent point (table I) and the immediate conclusion is

$$0 < \lambda_j, \ \lambda_{j+1}, \ldots, \lambda_n$$

Partly analogous to the method of Wilk et al. (ref. 3) some small number $m \geq 1$ of the smallest ordered mean squares are assumed to have been drawn from null populations so that testing begins with $j = m + 1$. Testing begins with the critical U_j values corresponding to a large nominal significance level α_p, where typically $0.25 \leq \alpha_p \leq 1.0$.

The special case of $\alpha_p = 1.0$ means that the number of items in the denominator remains fixed at $m + 1$, and testing proceeds at the final nominal level α_f, where typically $0.001 \leq \alpha_f \leq 0.05$. The test statistic for the j^{th} mean square is then

$$U_{m+1} = \frac{(m + 1)Z_j}{Z_1 + \ldots + Z_m + Z_j} \tag{9}$$

If $\alpha_p < 1.0$, the tests (with the use of eq. (8)) at nominal level α_p are called preliminary tests, and they continue so long as nonsignificance is the result. All Z_j testing nonsignificant remain in the denominator. This procedure is analogous to the familiar "sometimes pool" procedures that have been used in two-stage testing. Because α_p is large, early in the chain some Z_j would be selected as being too large to be an obvious member of a population with $\lambda_j = 0$. On the other hand, because α_p is large, there should be no great confidence that Z_j was drawn from a population with $\lambda_j > 0$. Therefore, a new level $\alpha_f < \alpha_p$ is imposed. If a Z_j is significant at level α_p, the same Z_j is tested at level α_f, also with the use of the U_j statistic. If Z_j is concluded to be significant at level α_f, then all $Z_k (k \geq j)$ are concluded to be significant.

TABLE I. - UPPER 100α PERCENT POINTS OF TEST STATISTIC U_j

Number of denominator mean squares, j	Nominal test size, α									
	0.001	0.002	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75
2	2.00000	1.99999	1.99997	1.99986	1.99917	1.99687	1.9877	1.923	1.706	1.382
3	2.9976	2.9960	2.9904	2.9809	2.951	2.904	2.806	2.527	2.086	1.688
4	3.976	3.962	3.925	3.870	3.760	3.625	3.412	2.949	2.395	1.961
5	4.887	4.845	4.758	4.65	4.44	4.21	3.89	3.287	2.658	2.184
6	5.74	5.63	5.46	5.31	4.99	4.68	4.28	3.57	2.893	2.371
7	6.51	6.33	6.11	5.87	5.46	5.09	4.61	3.83	3.11	2.54
8	7.20	6.96	6.65	6.35	5.88	5.44	4.91	4.06	3.29	2.69
9	7.81	7.52	7.10	6.78	6.26	5.75	5.17	4.27	3.45	2.82
10	8.34	8.01	7.53	7.17	6.59	6.03	5.41	4.45	3.60	2.95
11	8.82	8.44	7.95	7.53	6.89	6.28	5.61	4.62	3.74	3.07
12	9.26	8.84	8.33	7.87	7.13	6.50	5.81	4.77	3.87	3.17
13	9.67	9.21	8.68	8.16	7.37	6.71	5.99	4.92	3.99	3.27
14	10.05	9.55	8.95	8.42	7.59	6.91	6.15	5.05	4.10	3.37
15	10.40	9.86	9.20	8.66	7.79	7.07	6.30	5.17	4.20	3.46
16	10.72	10.14	9.43	8.83	7.96	7.23	6.44	5.29	4.30	3.55
17	11.01	10.40	9.64	9.00	8.12	7.38	6.57	5.40	4.39	3.63
18	11.28	10.64	9.84	9.17	8.28	7.52	6.69	5.50	4.48	3.70
19	11.53	10.86	10.03	9.34	8.43	7.65	6.81	5.60	4.56	3.77
20	11.76	11.07	10.22	9.51	8.58	7.78	6.92	5.69	4.64	3.84
21	11.98	11.28	10.40	9.67	8.72	7.90	7.03	5.78	4.71	3.90
22	12.19	11.48	10.58	9.83	8.86	8.02	7.13	5.87	4.78	3.96
23	12.39	11.68	10.76	9.99	8.99	8.13	7.23	5.95	4.85	4.02
24	12.58	11.87	10.93	10.14	9.12	8.24	7.33	6.03	4.92	4.08
25	12.76	12.05	11.10	10.29	9.23	8.34	7.42	6.11	4.98	4.14
26	12.93	12.22	11.26	10.43	9.34	8.44	7.51	6.18	5.04	4.19
27	13.09	12.38	11.41	10.56	9.44	8.54	7.60	6.25	5.10	4.24
28	13.24	12.53	11.55	10.68	9.54	8.63	7.68	6.32	5.16	4.30
29	13.39	12.68	11.68	10.78	9.64	8.72	7.76	6.38	5.22	4.35
30	13.53	12.82	11.80	10.88	9.74	8.81	7.83	6.44	5.28	4.40
31	13.67	12.96	11.91	10.98	9.83	8.89	7.90	6.50	5.33	4.45
32	13.80	13.09	12.01	11.07	9.91	8.97	7.97	6.56	5.38	4.50
33	13.93	13.21	12.10	11.16	9.99	9.04	8.04	6.62	5.43	4.54
34	14.05	13.32	12.19	11.25	10.07	9.11	8.11	6.68	5.48	4.58
35	14.17	13.43	12.27	11.34	10.15	9.18	8.17	6.74	5.53	4.62

TABLE I. - Concluded. UPPER 100α PERCENT POINTS OF TEST STATISTIC U_j

Number of denominator mean squares, j	Nominal test size, α									
	0.001	0.002	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75
36	14.29	13.53	12.35	11.43	10.22	9.25	8.23	6.80	5.58	4.66
37	14.41	13.63	12.43	11.51	10.29	9.31	8.29	6.85	5.63	4.70
38	14.53	13.73	12.51	11.59	10.36	9.37	8.35	6.90	5.67	4.74
39	14.64	13.82	12.59	11.67	10.43	9.43	8.41	6.95	5.71	4.78
40	14.75	13.91	12.67	11.75	10.50	9.49	8.46	6.99	5.75	4.82
41	14.85	14.00	12.75	11.83	10.57	9.55	8.51	7.03	5.79	4.86
42	14.95	14.09	12.83	11.90	10.64	9.61	8.56	7.07	5.83	4.90
43	15.05	14.17	12.90	11.97	10.70	9.67	8.61	7.11	5.87	4.94
44	15.15	14.25	12.97	12.04	10.76	9.72	8.66	7.15	5.91	4.98
45	15.24	14.33	13.05	12.11	10.82	9.77	8.71	7.19	5.95	5.01
46	15.33	14.40	13.12	12.18	10.88	9.82	8.76	7.23	5.99	5.04
47	15.42	14.47	13.19	12.25	10.94	9.87	8.81	7.27	6.03	5.07
48	15.50	14.54	13.26	12.32	11.00	9.92	8.85	7.31	6.07	5.10
49	15.58	14.60	13.32	12.38	11.06	9.97	8.89	7.35	6.11	5.13
50	15.66	14.66	13.38	12.44	11.11	10.02	8.93	7.39	6.14	5.16
51	15.73	14.72	13.44	12.50	11.16	10.07	8.97	7.43	6.17	5.19
52	15.80	14.79	13.50	12.56	11.21	10.12	9.01	7.47	6.20	5.22
53	15.87	14.85	13.56	12.62	11.26	10.17	9.05	7.51	6.23	5.25
54	15.93	14.91	13.62	12.68	11.31	10.21	9.09	7.55	6.26	5.28
55	15.99	14.97	13.67	12.73	11.36	10.25	9.13	7.59	6.29	5.31
56	16.05	15.03	13.72	12.78	11.40	10.29	9.17	7.63	6.32	5.34
57	16.11	15.10	13.77	12.83	11.44	10.33	9.21	7.67	6.35	5.37
58	16.17	15.16	13.82	12.88	11.48	10.37	9.25	7.70	6.38	5.40
59	16.23	15.22	13.87	12.93	11.52	10.41	9.29	7.73	6.41	5.43
60	16.29	15.28	13.92	12.97	11.56	10.45	9.33	7.76	6.44	5.46
61	16.34	15.34	13.97	13.01	11.60	10.49	9.37	7.79	6.47	5.48
62	16.39	15.40	14.02	13.05	11.64	10.53	9.41	7.82	6.50	5.50
63	16.44	15.46	14.06	13.09	11.67	10.57	9.45	7.85	6.53	5.52

Suppose Z_j is significant at level α_p, but not at level α_f. Continue testing Z_{j+1}, . . . until some Z_k is reached that tests significant at α_f, where only the first j - 1 mean squares were judged to be from null populations:

$$U_j = \frac{jZ_k}{Z_1 + \ldots + Z_{j-1} + Z_k} \tag{10}$$

If some Z_k is significant at level α_f, all larger mean squares are also concluded to be significant.

Computation of U_j Distribution

The chain pooling procedures require nominal significance levels additional to those given in reference 8. Critical points of U_j were obtained with the use of a Monte

TABLE II. - COMPARISON OF PERCENTAGE POINTS OF STATISTIC U_j OBTAINED FROM MONTE CARLO COMPUTATIONS WITH PERCENTAGE POINTS OF U_j COMPUTED FROM TABLES OF COCHRAN'S STATISTIC[a]

Number of denominator mean squares, j	Nominal test size, α					
	0.01			0.05		
	Critical point, U_j		Difference	Test statistic, U_j		Difference
	Cochran	Monte Carlo		Cochran	Monte Carlo	
2	1.9998	1.99986	0.0001	1.9970	1.99687	-0.0001
3	2.9799	2.9809	.0010	2.9007	2.904	.003
4	3.8704	3.870	.000	3.6260	3.625	-.001
5	4.6395	4.65	.01	4.2060	4.21	.00
6	5.2968	5.31	.01	4.6848	4.68	.00
7	5.8632	5.87	.01	5.0897	5.09	.00
8	6.3560	6.35	-.01	5.4384	5.44	.00
9	6.7896	6.78	-.01	5.7465	5.75	.00
10	7.1750	7.17	-.005	6.0200	6.03	.01
12	7.8336	7.87	.04	6.4920	6.50	.01
15	8.6205	8.66	.04	7.0635	7.07	.01
20	9.5980	9.51	-.09	7.7880	7.78	-.01
24	10.1928	10.14	-.05	8.2416	8.24	.00
30	10.8960	10.88	-.02	8.7870	8.81	.02
40	11.7600	11.75	-.01	9.4800	9.49	.01
60	12.9060	12.97	.06	10.4220	10.45	.03

[a]Ref. 8.

Carlo method, as described in reference 9, and are given in table I. These points are compared with Cochran's statistic in table II, and the comparison suggests that the results are in agreement to better than one unit in the next to the last figure reported.

EVALUATION OF CHAIN POOLING PROCEDURES

Risk Curves

The relative importance of type 1 and type 2 errors depends on the application so that a general evaluation of a decision procedure cannot combine the type 1 and type 2 losses. A procedure is usually chosen that involves nonzero probabilities of both types of errors, but the losses associated with one or the other type may dictate that a strong effort be made to control just one of the two types. A compromise is chosen where a lowered probability of one type of error will increase the probability of the other. An evaluation of the operating characteristics of a multiple decision procedure must therefore exhibit the set of possible compromises.

The operating characteristics of the chain pooling procedures will be determined by Monte Carlo methods. The results, therefore, will show the relative frequency with which type 1 and type 2 decision errors have been made. Multiplying the observed relative frequencies by the type 1 and type 2 losses, as given in equations (4) and (7), produces quantities that will be called the observed type 1 and type 2 risks.

For a given procedure, a set of choices (such as differing values of nominal α) could produce a set of pairs of risk values R_1 and R_2 as indicated, for a hypothetical

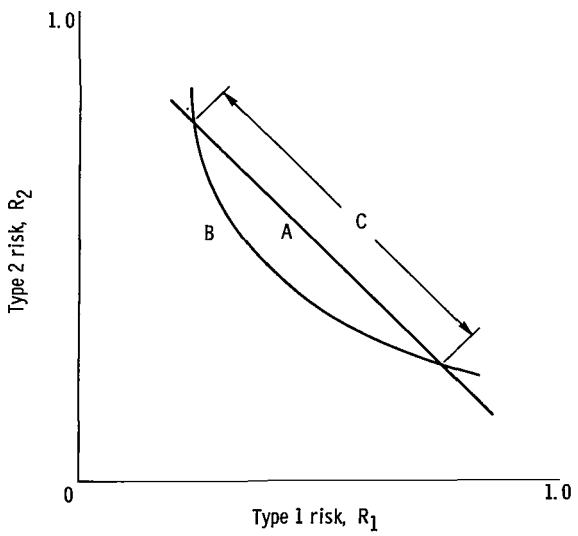


Figure 1. - Considerations in evaluation of pooling procedures.

example, by the set of points consisting of curve A of figure 1. On the other hand, a different procedure could lead to a set of risk values such as those of curve B. Over the range C for any given R_1, curve B has lower R_2 than curve A or for any given R_2, curve B has lower R_1 than curve A. The procedure resulting in curve B, therefore, is a preferred procedure over the range C with respect to the procedure that gave curve A.

Parameters of Risk Curves

A digital computer (IBM 7094) was used to generate pseudonormal random variables, as described in reference 9. For 2^ℓ treatments, 2^ℓ such numbers were used with Yates' algorithm to compute mean squares. The use of Yates' algorithm with pseudonormal variates gives contrasts that are the sum of 2^ℓ approximately normal variates. The central limit theorem therefore implies that an improved approximation to normality was obtained, over what would have been obtained if pseudo $\sigma^2\chi^2_{(1)}$ variates had been generated more directly.

Suppose that the contrasts computed with Yates' algorithm are listed in their order of computation. As in typical experiments with real data, the first mean square for total or grand mean is excluded from further consideration. Each subsequent contrast is augmented by the addition of an increment, $2^\ell\delta_i\theta\sigma$, for $i = 1, \ldots, \rho$; $\rho \leq 2^\ell - 2$. In terms of equation (5), the increment is $2^\ell\mu_i$; that is, the parameter estimated by the contrast has been given the value μ_i. The mean squares, therefore, have noncentrality parameters, as given by equation (3).

In the analysis of a physical experiment, the two types of parameters are the unknown parameters of the populations as chosen by "nature" and the parameters of the ANOVA strategy as chosen by the statistician. The unknown parameters are ρ, σ^2, and λ_i; $i = 1, \ldots, \rho$. The parameters assigned by the statistician in the case of chain pooling are m, α_p, and α_f. The generation of Monte Carlo experiments requires that values must also be assigned to ρ, σ^2, and λ_i. Because the procedure is scale invariant, the investigation can be simplified to $\sigma^2 = 1$.

The values of ρ and λ_i are to be chosen so that they will impose a severe burden on the available strategies. Application of a wide variety of strategies should then provide a demarcation of the superior strategies. The strategy evaluations should be based on combinations of λ_i values that are especially unfavorable with respect to type 2 errors. A complete investigation of the operating characteristics of the multiple decision procedures would evaluate ρ type 2 error probabilities as joint functions of the $\lambda_1, \lambda_2, \ldots, \lambda_\rho$. Such an investigation would not be readily interpretable. The problem is simplified by defining a type 2 risk as the expectation, over the experiment,

of losses due to type 2 errors. Obviously, the type 2 risk will be sensitive to the distribution of the λ_i over $i = 1, \ldots, \rho$. The scope of the investigation is reduced to manageable size by considering only such distributions of the λ_i as should result in especially unfavorable operating characteristics.

Unfavorable Distribution of Parameters

If all the Z_i had come from one central $\sigma_0^2\chi^2_{(1)}$ distribution, the correct decisions would be that all $\lambda_i = 0$. The order statistics Z_j would have some set of expectations $\sigma_0^2\zeta_j$. On the other hand, suppose that the Z_i were drawn from n noncentral chi-square populations, each having a noncentrality parameter such that

$$(1 + \lambda_i)\sigma^2 = \sigma_0^2\zeta_j;\ i, j = 1, \ldots, n$$

The resulting mean squares would thus have expectations equal to the expectations of the order statistics of the central $\sigma_0^2\chi^2_{(1)}$ distribution. This equating of expectations of mean squares from n noncentral populations to the expectations of the ordered observations of the single central population implies that, for the noncentral populations, the multiple decision procedure might have selection probabilities (power functions) no larger than the type 1 error probabilities for the central $\sigma_0\chi^2_{(1)}$ population. Such a set of λ_i values would therefore constitute a distribution that would be unfavorable to the selection probabilities under chain pooling.

From equations (3) and (5),

$$\lambda_i = \frac{2^\ell \mu_i^2}{\sigma^2} = 2^\ell \theta^2 \delta_i^2$$

Therefore, an unfavorable distribution of the λ_i can be obtained by setting

$$\delta_i^2 = \zeta_{\rho-j+1}$$

where the ζ_j are the expectations of the order statistics of a sample of size ρ from the central $\chi^2_{(1)}$ distribution. (These values of ζ_j are such that $\delta_i^2 = \zeta_{\rho-j+1}$ will satisfy eq. (6)). Expectations of order statistics from a gamma distribution with

scale parameter 1, shape parameter 1/2, and many sample sizes were tabulated in reference 10. Multiplying these values by 2 gives the expectations of the order statistics of the central $\chi^2_{(1)}$ distribution. The Monte Carlo experiments were performed with such unfavorable sets of λ_i values except that the number ρ of $\lambda_i \neq 0$ was less than n.

In general, efficient experimentation is achieved when η is small in comparison with ρ, and this is achieved when the experimenter uses his prior knowledge to choose that fractional factorial design that results in most of the mean squares being significant. Correspondingly, values of η of 4, 6, and 9 were investigated for 2^4 treatment combinations; values of 6, 9, and 14 were investigated for the 2^5 case; and values of 13, 23, and 33 were investigated for the 2^6 case. All the distributions of the λ_i were obtained from reference 10 for values of $\rho = 6$ to $\rho = 40$. For the single case of $\rho = 50$, the distribution of λ_i was obtained with the use of an approximation to ζ_j as described in reference 9.

The fixing of the distribution of the λ_i allows the type 2 risk to be investigated as a function of a single parameter λ, where λ is the mean of the λ_i over $i = 1, \ldots, \rho$:

$$\lambda = \frac{1}{\rho} \sum_{i=1}^{\rho} \lambda_i \tag{11}$$

From equations (3), (5), (6), and (11),

$$\lambda = \frac{1}{\rho} \sum_{i=1}^{\rho} \frac{2^{\ell} \mu_i^2}{\sigma^2}$$

$$= \frac{1}{\rho} \sum_{i=1}^{\rho} 2^{\ell} \theta^2 \delta_i^2$$

$$= 2^{\ell} \theta^2 \tag{12}$$

Minimax and Bayes Strategies

Monte Carlo investigations to select from many strategies (m, α_p, α_f) the strategy best against a very unfavorable distribution of the λ_i should result in what might be called an "empirical minimax procedure." The problem may also be approached from a Bayesian point of view. It would require the assumption of a prior probability dis-

tribution of the effects μ_i. The μ_i in an experimental situation might be the additive result of several diverse sources. The μ_i could then be regarded as a sample of size ρ from an approximate prior normal distribution with mean zero, and in this situation it was assumed that nature is a disinterested opponent. The δ_j^2 would then be the order statistics of a sample of size ρ from a $\chi^2_{(1)}$ distribution. In other words, if nature is a disinterested opponent, the same prior distribution of the δ_i^2 should occur as would be anticipated from an aggressive opponent. The Monte Carlo founded minimax procedure is then also the Monte Carlo founded Bayes' procedure.

Scorekeeping for Monte Carlo Experiments

After being augmented, the Z_i are ordered in ascending rank, and the m smallest Z_i are presumed to have been drawn from null populations. The next $2^\ell - 1 - m$ mean squares are examined for significance in accordance with equations (9) or (10). A type 1 error is counted for the test of Z_i if both (1) the test of Z_i resulted in rejection of the null hypothesis, and (2) the particular Z_i is a mean square that had not been augmented. A type 2 error is counted for the test of Z_i if both (1) the test of Z_i resulted in acceptance of the null hypothesis, and (2) the particular Z_i is a mean square that had been augmented with $\delta_i > 0$. In this way, N experiments are analyzed, each containing ρ violations of the null hypothesis. In all cases, N = 1000 and for given ℓ, the same $(1000)2^\ell$ pseudonormal variates were used for every strategy investigated.

The mean squares in Yates' order from $i = \rho + 1$ to $i = 2^\ell - 1$ were not augmented. For these mean squares, and over the N experiments, the computer counts the number of type 1 errors and divides by N to report the observed type 1 error probability P_{1i} for the i^{th} mean square. The number of P_{1i} computations in any experiment is $\eta = 2^\ell - \rho - 1$; however, not all $\eta + \rho$ mean squares were tested. The m smallest mean squares were pooled before testing, and, therefore, only $\eta - m$ opportunities should be expected for making type 1 errors.

As given in equation (4), the type 1 errors are defined as unit losses. The observed risk (av type 1 loss) is estimated by P_{1i} as averaged over the experiment:

$$\bar{R}_1(\eta) = \frac{1}{\eta - m} \sum_{i=\rho+1}^{2^\ell - 1} P_{1i} = \frac{1}{2^\ell - \rho - m - 1} \sum_{i=\rho+1}^{2^\ell - 1} P_{1i} \qquad (13)$$

The symbol η is attached to $\bar{R}_1$ to represent a fact that will be developed later; that is, $\bar{R}_1$ is mainly a function of η.

Equation (13) does not provide a strict probability. In any experiment, random fluctuations can cause some of the ρ augmented mean squares to have smaller values than some of the η mean squares not so augmented. If the strategy used a large value of m, some of the smallest augmented mean squares could be pooled into the initial denominator of the test statistic, and more than η - m of the null mean squares could be available for type 1 errors, which could result in $\overline{R}_1(\eta)$ being greater than 1.

The first ρ mean squares beyond the grand mean were augmented so that $\lambda_i > 0$. The number of type 2 errors for each $i = 1, \ldots, \rho$ is counted over the N experiments and divided by N to report the observed type 2 error probability P_{2i} for the i^{th} augmented mean square. Multiplication of these probabilities by the losses, as given in equation (7), allows the observed risk to be computed over the experiment. Thus,

$$\overline{R}_2(\lambda) = \frac{1}{\rho} \sum_{i=1}^{\rho} \delta_i^2 P_{2i} \tag{14}$$

The symbol λ is attached to $\overline{R}_2$ to indicate a fact that will be developed later; namely, $\overline{R}_2$ is mainly a function of λ.

Details of the decision and scorekeeping procedure as it was built into a computer program are given in reference 9.

CHOICE OF PRELIMINARY TEST LEVEL

Conditions of m, ρ, and λ

Those values of α_p that should be preferred for a wide variety of conditions will now be evaluated for several arbitrarily chosen values of m. A later step will determine preferred values of m, given that the α_p are already preferred values.

The values of ρ will be much larger than the value of $\rho = 1$ in Daniel's investigation (ref. 2). (Comparisons of chain pooling with the operating characteristics of procedures based on half-normal plotting were presented in ref. 9, in which the chain pooling procedures were shown to be superior to the results obtained by Daniel and by Birnbaum.)

Values of λ were chosen to result in $\overline{R}_2(\lambda)$ values that cover the range of $0.05 \leq R_2(\lambda) \leq 0.20$.

Experiments of 2^4 Treatment Combinations

Some results for $\ell = 4$, $\rho = 6$, $\eta = 9$, and $m = 1$ are shown in figure 2. The strategies consisted of $m = 1$ together with the values of α_p identified by the symbols

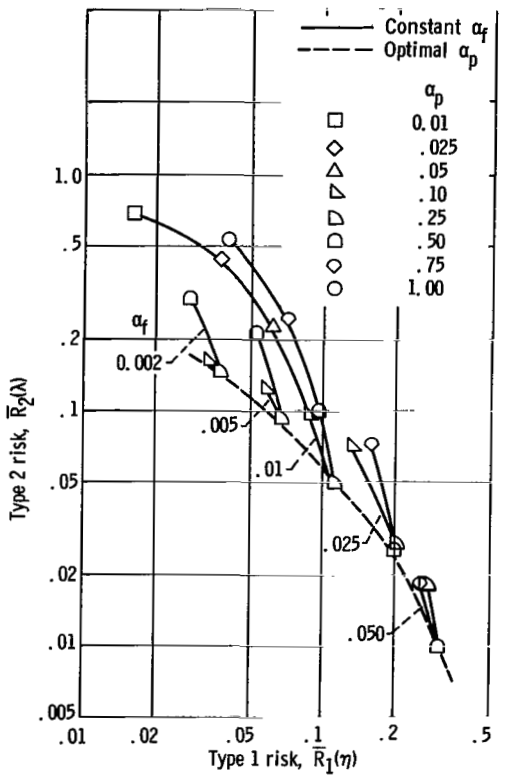


Figure 2. - Risk curves. $\ell = 4$, $\rho = 6$, $\eta = 9$, $\lambda = 64$, $m = 1$.

and the values of α_f that identify the solid curves. The set of preferred strategies (the set of points nearest the origin) is the set that jointly minimizes $\bar{R}_1(\eta)$ and $\bar{R}_2(\lambda)$. These points are identified by the dashed curve drawn through them, and they include $\alpha_p = 0.25$ at the smaller values of $\bar{R}_1(\eta)$, and $\alpha_p = 0.50$ at the larger values of $\bar{R}_1(\eta)$. Similar results for $m = 2$ and $m = 3$ are presented in reference 9.

An important implication of figure 2 is that a single value of α_p cannot be preferred for all values of α_f. However, on scanning the values of α_p that are identified as being preferred by their lying on the dashed curve, selections can be made of those values of α_p that should be preferred for given values of the abscissa $\bar{R}_1(\eta)$. For example, figure 2 shows that the vertical line at $\bar{R}_1(\eta) = 0.05$ cuts the dashed curve between two points for which $\alpha_p = 0.25$, while at $\bar{R}_1(\eta) = 0.30$, $\alpha_p = 0.50$ would be preferred. In some cases (such as $\bar{R}_1(\eta) = 0.20$ of fig. 2) there is no clear choice

TABLE III. - PREFERRED NOMINAL SIZE OF PRELIMINARY POOLING TEST α_p

(a) 2^4 Treatment combinations

Number of mean squares pooled before testing, m	Average noncentrality parameter, λ	Type 1 risk, $\overline{R}_1(\eta)$	Number of null mean squares, η: 4	6	9
			Nominal size of preliminary pooling test, α_p		
1	64	0.05	[a]0.25 0.50	0.25	0.25
	81	↓	[a].25 .50	.25	----
	100	↓	[a].25 .50	.25	----
	64	.10	.50	.25, 0.50	.25
	81	↓	.50	.25, .50	.25
	100	↓	.50	.25, .50	.25
	64	.20	.50	.50	.25, [a]0.50
	81	↓	.50	.50	.25, [a].50
	100	↓	.50	.50	.25, [a].50
2	64	0.05	[a]0.50 0.75	0.50	0.50
		.10	.75	.50 [a]0.75	.50 [a]0.75
		.20	[a].75 1.00	.75	.75
3	64	0.05	0.75	0.75	0.75
		.10	[a].75 1.00	[a].75 1.00	.75
		.20	1.00	1.00	.75 [a]1.00

(b) 2^5 Treatment combinations

Number of mean squares pooled before testing, m	Average noncentrality parameter, λ	Type 1 risk, $\overline{R}_1(\eta)$	Number of null mean squares, η: 6	9	14
			Nominal size of preliminary pooling test, α_p		
1	64	0.05	0.50	0.50	0.25
	64	.10	.50	↓	.25
	81	.10	----	↓	----
	49	.20	----	↓	.50
	64	↓	.50	↓	.50
	81	↓	----	↓	----
3	16	0.05	0.75	----	----
	64	.05	.75	0.75	0.75
	16	.10	.75	----	----
	64	.10	.75	.75	.75
	16	.20	1.00	----	----
	64	.20	1.00	1.00	.75 [a]1.00
5	64	0.05	1.00	1.00	1.00
	16	.10	↓	----	----
	64	.10	↓	1.00	1.00
	16	.20	↓	----	----
	64	.20	↓	1.00	1.00

[a]Use of this value makes α_p independent of η at given m.

between two values of α_p ($\alpha_p = 0.25$ and $\alpha_p = 0.50$ are equally preferred). Results similar to those for figure 2 are shown for $\eta = 6$ and for $\eta = 4$ in reference 9.

The preferred or equally preferred values of α_p, as obtained for the conditions detailed in reference 9, are summarized in table III(a). This summary shows that the preferred values of α_p are completely independent of the noncentrality parameter λ and that the preferred α_p can be selected to be independent of the number of null effects η.

Experiments of 2^5 Treatment Combinations

The type of evaluation of chain pooling strategies just described for the case of $\ell = 4$ was also carried out for $\ell = 5$. Results are presented in reference 9. The implications of these results for preferred values of α_p are given in table III(b). This table shows again that the preferred values of α_p are independent of the average noncentrality parameter λ and that they are almost independent of the number of null mean squares η.

Experiments of 2^6 Treatment Combinations

Chain pooling strategies for $\ell = 6$ were investigated at values of $\rho = 50$, 40, and 30, and the corresponding values of η were 13, 23, and 33. Results with $m = 1$ and two values of λ are shown by figures 3(a), 3(b), and 3(c), respectively. The results for $\eta = 13$ and for $\eta = 23$ (figs. 3(a) and 3(b)) show that with $m = 1$, the preferred value of α_p is 1.0, as shown by the dashed line. A preferred value of $\alpha_p = 1.0$ is also shown with $\eta = 33$ by figure 3(c) for $\lambda = 64$, but figure 3(c) shows that other values of α_p are preferred when $\lambda = 256$. The desirability of using small values of α_p, shown by figure 3(c) at large η and λ is a nontypical phenomenon discussed in detail in reference 9. In brief, if, first, the value of η is large relative to ρ, and if, second, η is large on an absolute basis, and if, third, λ is large enough to ensure low type 2 losses, the preferred value of α_p can be quite small. Such a favorable combination of ρ, η, and λ would not ordinarily be known to exist, a priori, and the use of small values of α_p cannot be regarded as a generally useful strategy. However, if the Monte Carlo investigations show that two values of α_p are equally preferred for the especially unfavorable distributions of δ_i, the smaller of the two α_p values could be used for the ANOVA of real experiments on the chance that the δ_i might have a more favorable distribution. (Some equally preferred pairs of α_p values are listed in table III.)

The use of values of $m > 1$ with $\ell = 6$ is illustrated by figure 4 for $m = 3$. These

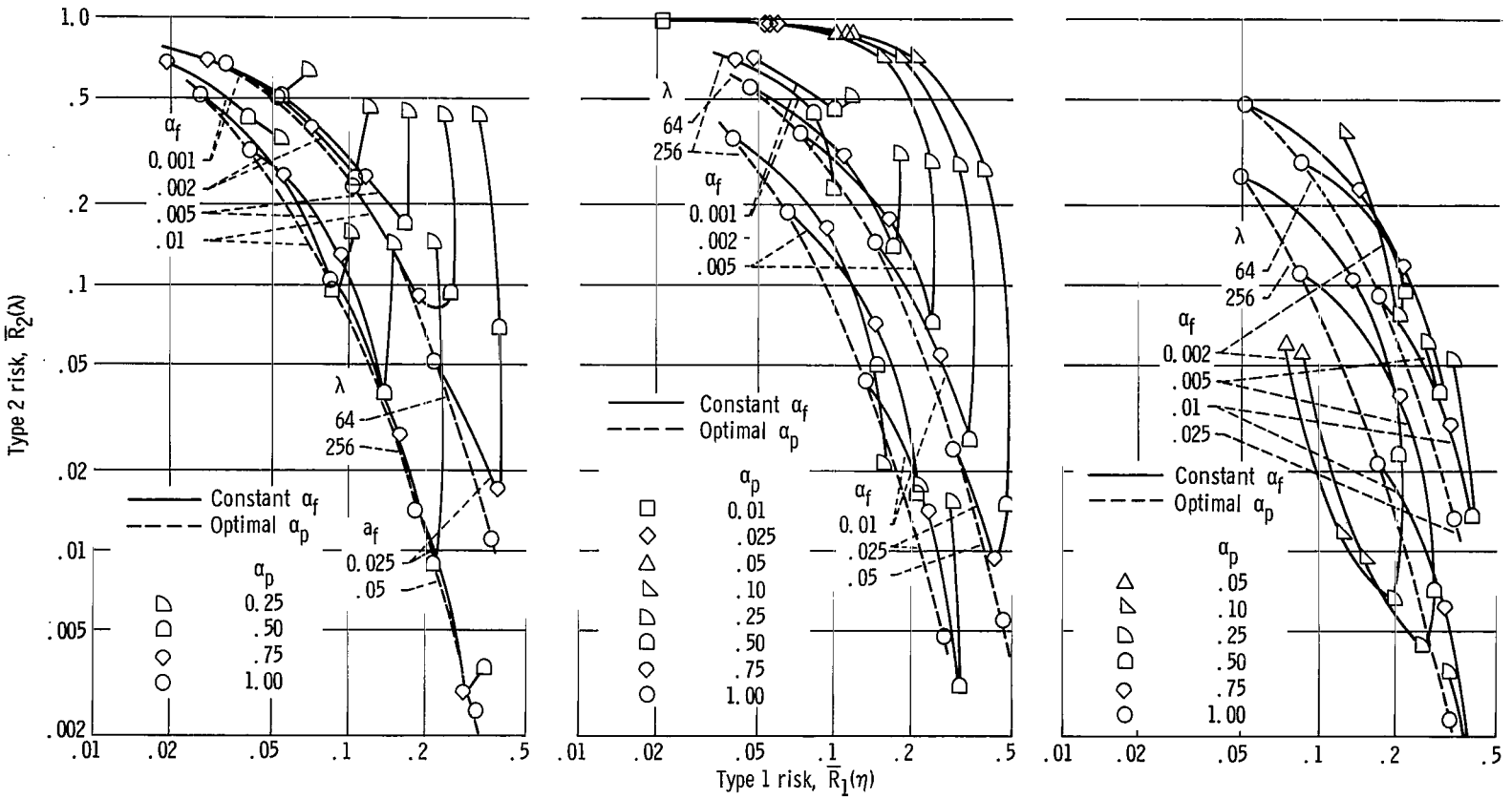


Figure 3. - Risk curves. $\ell = 6$ and $m = 1$.

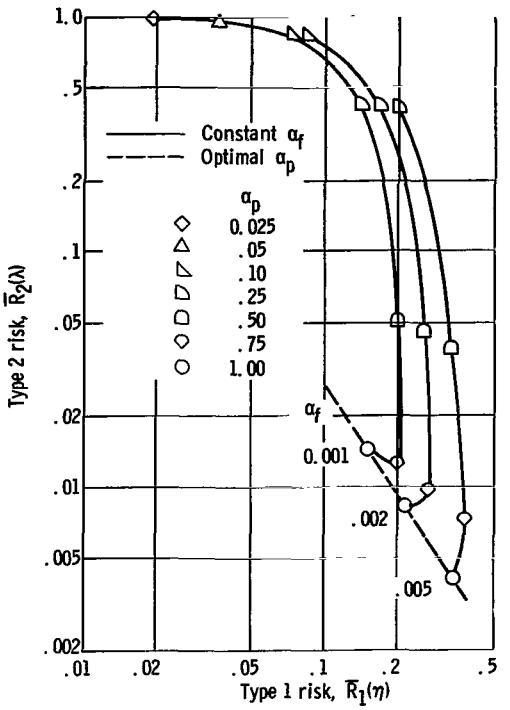


Figure 4. - Risk curves. $\ell = 6$, $\lambda = 64$, $\rho = 40$, $\eta = 23$, and $m = 3$.

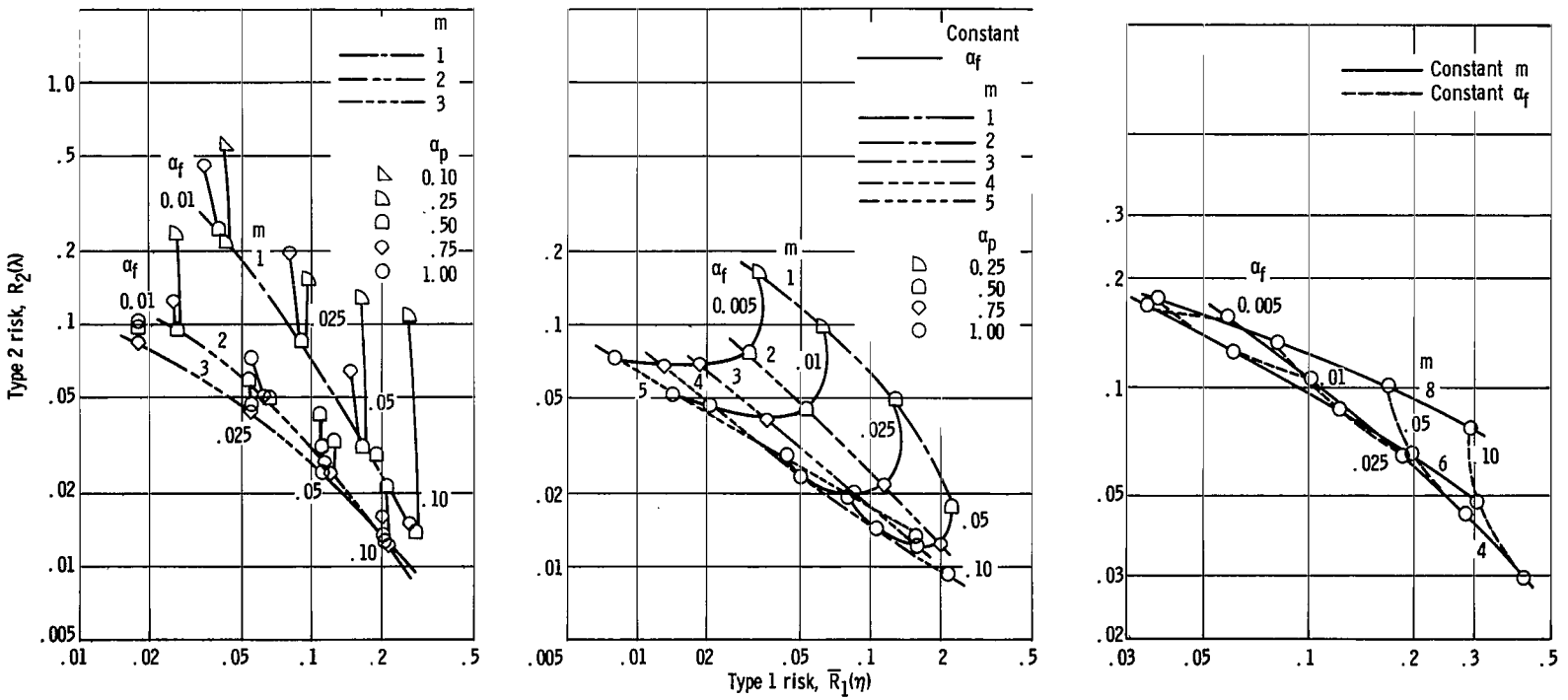
results show that the preferred value of α_p is 1.0, but that for the value of $\lambda = 64$ (which gives reasonable values of $\overline{R}_2(\lambda)$) there were no values of α_f small enough to give values of $\overline{R}_1(\eta)$ as small as might reasonably be desired.

NUMBER OF MEAN SQUARES INITIALLY POOLED

The results that give preferred values of α_p for arbitrarily chosen values of m have been presented. Now, the results that will determine the preferred values of m are presented for values of α_p that are already preferred.

Experiments of 2^4 Treatment Combinations

Some risk curves with $\eta = 4$ are presented in reference 9. These results showed preferred values of α_p (which were independent of λ) for $m = 1$, 2, and 3. These results are displayed in figure 5(a) but only for the single value of the noncentrality parameter $\lambda = 64$ (which gives reasonable values of $\overline{R}_2(\lambda)$). The curves drawn through the preferred α_p points of figure 5(a) show that $m = 2$ or $m = 3$ is greatly preferred over $m = 1$, but that $m = 3$ is only a slight improvement over $m = 2$.

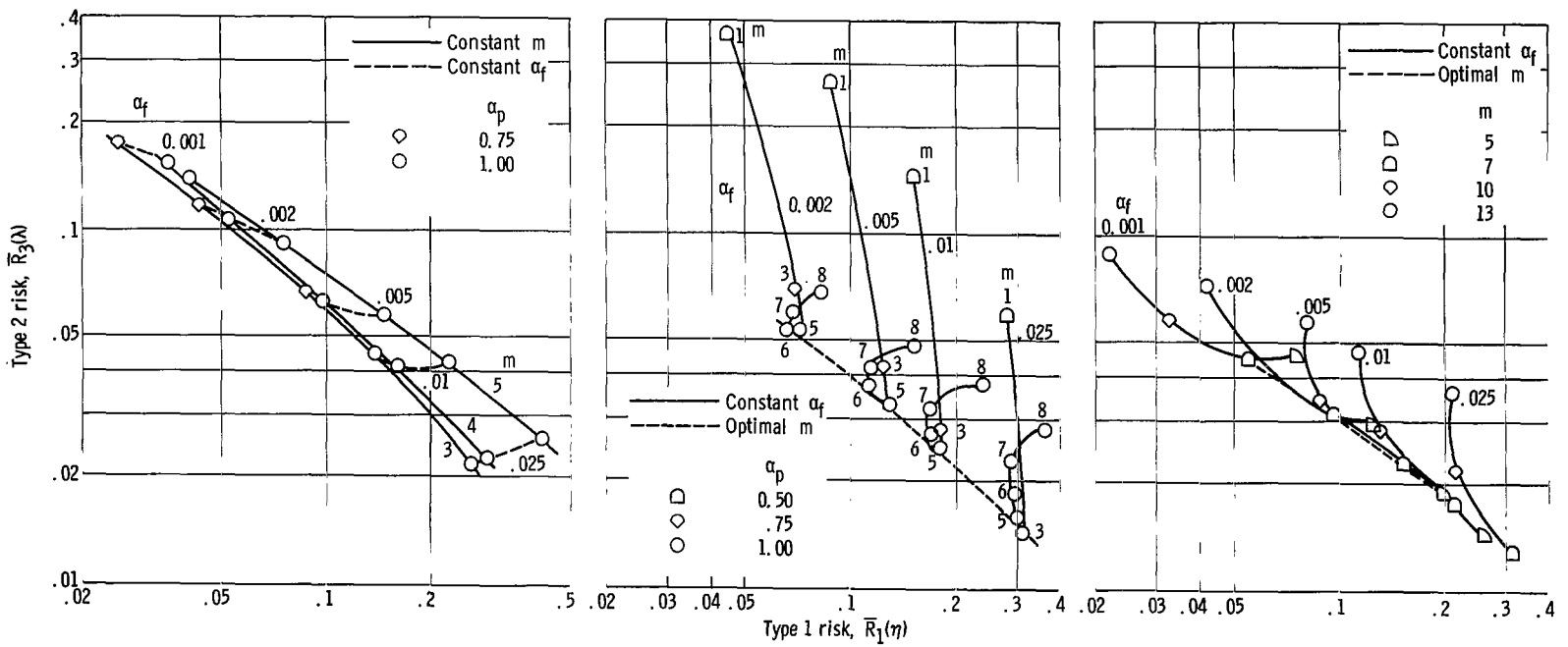


(a) $\rho = 11$, $\eta = 4$, $\lambda = 64$.

(b) $\rho = 9$, $\eta = 6$, $\lambda = 64$.

(c) $\rho = 6$, $\eta = 9$, $\lambda = 16$.

Figure 5. - Risk curves. $\ell = 4$.



(a) $\rho = 25$, $\eta = 6$.

(b) $\rho = 22$, $\eta = 9$.

(c) $\rho = 17$, $\eta = 14$, $\alpha_p = 1.0$.

Figure 6. - Risk curves. $\ell = 5$, $\lambda = 32$.

At a larger value of η, the risk curves for $\eta = 6$ are shown in figure 5(b) for preferred values of α_p and for values of m of 1, 2, 3, 4, and 5. The strategy of $m = 4$ is seen to be preferred over the important range of $0.05 \leq \overline{R}_1(\eta) \leq 0.20$.

Results for $\eta = 9$ are shown in figure 5(c) for preferred values of α_p, and for values of m of 4, 6, and 8. The strategy of $m = 6$ is seen to be preferred over the important range of $0.05 \leq \overline{R}_1(\eta) \leq 0.20$.

Experiments of 2^5 Treatment Combinations

Results with 2^5 treatments for $\lambda = 32$, $\eta = 6$, and $m = 3$, 4, and 5, are shown in figure 6(a) for preferred values of α_p. In the important range of $0.05 \leq \overline{R}_1(\eta) \leq 0.20$, these results show that $m = 3$ is the preferred strategy. Similar results for $\eta = 9$ are shown in figure 6(b), and these results show $m = 5$ or $m = 6$ to be a preferred strategy. For $\eta = 14$, figure 6(c) shows that $m = 7$ is the best strategy over the range of $0.05 \leq \overline{R}_1(\eta) \leq 0.20$.

Experiments of 2^6 Treatment Combinations

The results achieved with $\ell = 6$ and with $m = 1$ and 3 (figs. 3 and 4) showed that $\alpha_p = 1.00$ was a preferred strategy with $m = 1$ but, that for $\eta = 23$, there were no suitable α_p and α_f values with $m = 3$. The large value of $\eta = 23$ suggests that suitable values of α_p and α_f might be found at values of m much larger than $m = 3$. Results for $\lambda = 64$ and for $\alpha_p = 1.00$ and 0.75 are shown for a wide variety of values of m by figure 7(a). These computations were performed by using the smallest available value of α_f, namely, $\alpha_f = 0.001$, because the results in figure 4 show that achieving values of $\overline{R}_1(\eta)$ in the (desired) range down to 0.05 would be very difficult with $m > 1$.

The points displayed in figure 7(a) show that the strategies with $\alpha_p = 1.00$ are better than those with $\alpha_p = 0.75$.

With the use of the preferred value of $\alpha_p = 1.00$, as indicated by figure 7(a), and a preferred set of m values as suggested by the lower branch of the curve of figure 7(a) (namely, $m = 5$, 10, 15, 20, and 21), the question of preferred α_f values was examined for the conditions of figure 7(b). The results presented in figure 7(b) show that, in any attempt to control $\overline{R}_1(\eta)$ in the range of 0.05 to 0.20, the preferred strategy consists of using α_f at its smallest available value (namely, $\alpha_f = 0.001$). The control of $\overline{R}_1(\eta)$ is then accomplished by the selection of a suitable value of m.

The case of $\rho = 50$ and $\eta = 13$ was examined for $\alpha_p = 1.0$, $\alpha_f = 0.001$, and a

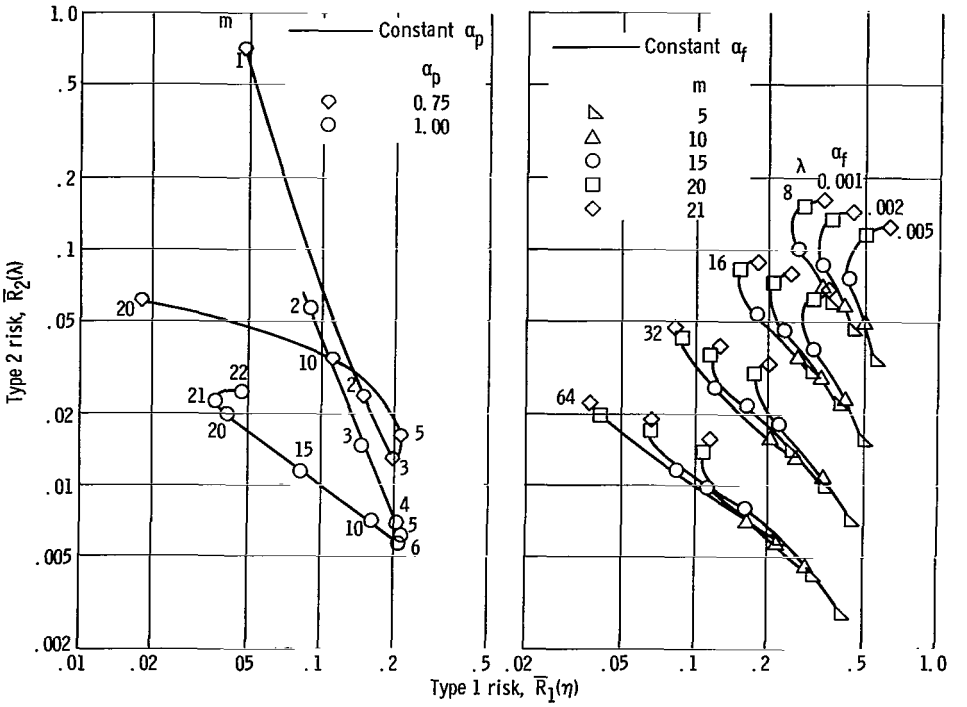
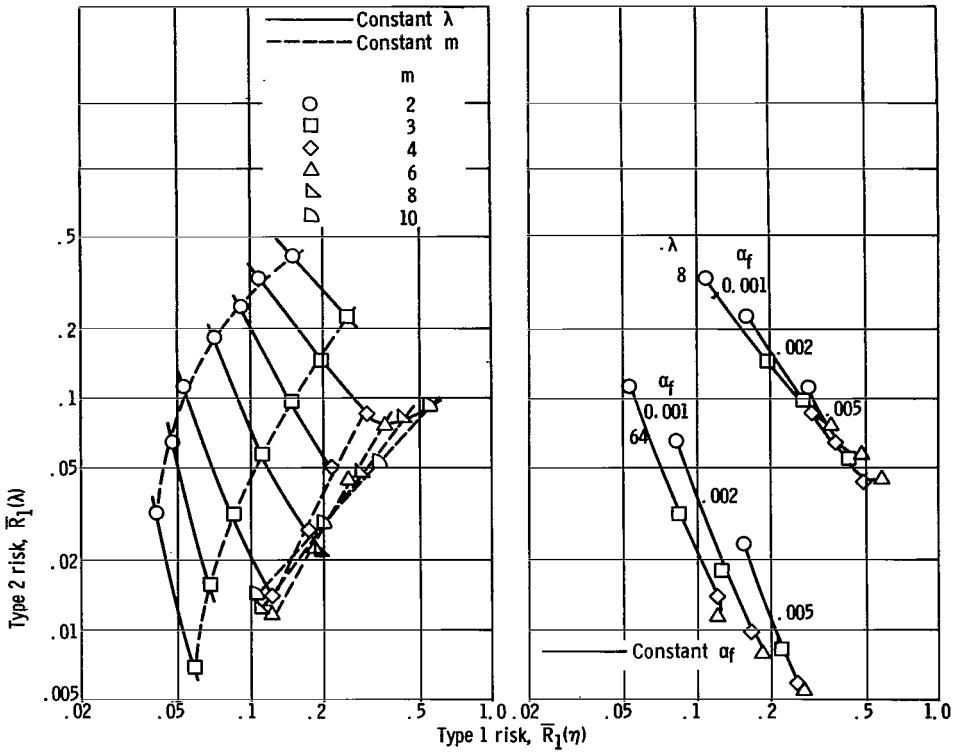


Figure 7. - Risk curves. $\ell = 6$, $\rho = 40$, and $\eta = 23$.

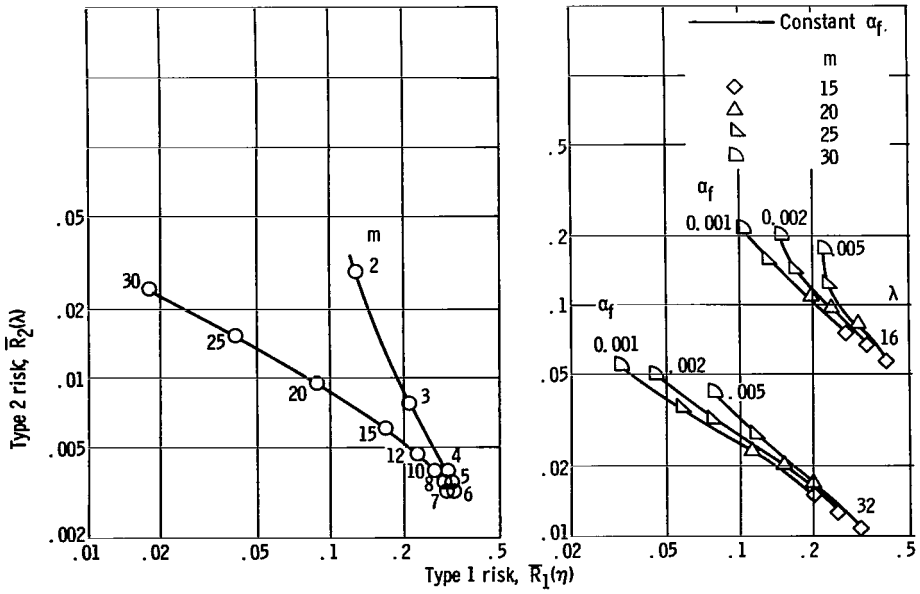
wide variety of values of m as shown by figure 8(a). These results show that m = 2, 3, 4, and 6 constitute an attractive set of m values. The selection of preferred α_f values for such a set of m values was examined with results shown in figure 8(b). These results are consistent with those shown in figure 7(b) for $\eta = 23$; namely, the strategy of selecting α_f at its smallest value ($\alpha_f = 0.001$) and controlling $\overline{R}_1(\eta)$ through the choice of m dominates the strategy of fixing m and controlling $\overline{R}_1(\eta)$ by selecting an α_f.

Results for $\eta = 33$ are shown in figure 9(a) for $\alpha_p = 1.0$, $\alpha_f = 0.001$, and a wide variety of values of m. The results are similar to those of figures 7(a) and 8(a) in that the risk curve has a turning point at a critical value of m. For values of m in excess of the critical value, there is a sequence of m values along a lower branch of the loss curve, and such a sequence constitutes a set of preferred m values, where increasing the value of m within the preferred set reduces the value of $\overline{R}_1(\eta)$ at the cost of increased $\overline{R}_2(\lambda)$. Figure 9 shows that setting α_f at 0.001 and controlling $\overline{R}_1(\eta)$ through m is a preferred strategy for $\eta = 33$, which is consistent with the results already cited for $\eta = 13$ and $\eta = 23$.



(a) $\alpha_p = 1.0$, $\alpha_f = 0.001$. (b) $\alpha_p = 1.0$

Figure 8. - Risk curves. $\ell = 6$, $\rho = 50$, and $\eta = 13$.



(a) $\lambda = 64$, $\alpha_f = 0.001$. (b) $\lambda = 16, 32$; $\alpha_f = 0.001, 0.002, 0.005$.

Figure 9. - Risk curve. $\ell = 6$, $\rho = 30$, $\eta = 33$, and $\alpha_p = 1.0$.

From inspection of figures 7(a), 8(a), and 9(a), the preferred ranges of m are presented in the following table:

Number of null mean squares, η	Preferred range of number of mean squares pooled before testing, m
13	2 to 6
23	6 to 21
33	6 to 30

The total results for 2^6 design experiments suggest that acceptable strategies consist of using $m = 1$, $\alpha_p = 1.0$ and α_f values small enough to give satisfactory values of $\overline{R}_1(\eta)$. Values of $m > 1$ can be used, but if they are, then according to the preceding table they should be at least as large as about $(1/5)\eta$ and then $\alpha_p = 1.0$, $\alpha_f = 0.001$, and values of m large enough to give satisfactory values of $\overline{R}_1(\eta)$ should be used.

Overall Procedure

A reasonable strategy for the chain pooling ANOVA of a 2^ℓ design experiment where $\ell = 4$ or 5 might consist of performing a preliminary analysis with $m = 1$ and whatever α_p value is shown by existing results (table III) to be preferred for the statistician's prior guess of the value of η. (For $\ell = 6$ and $m = 1$, the preferred α_p is equal to 1.0.) From the results of the chain pooling ANOVA with $m = 1$,

TABLE IV. - RATIO OF m/η FOR GIVEN η AND PREFERRED m

Number of treatment combinations, 2^ℓ	Number of null mean squares, η	Number of mean squares pooled before testing, m	m/η
2^4	4	3	0.75
2^4	6	4	.67
2^4	9	6	.67
2^5	6	3	.50
2^5	9	5	.56
2^5	14	7	.50

the statistician will have a posterior estimate $\hat{\eta}$ of η and should use this value of $\hat{\eta}$ to pick a preferred value of m that will be greater than 1.

A rule is needed for picking a value of $m > 1$ after $\hat{\eta}$ has been determined. The preferred values of m, as evidenced by figures 5 and 6 for $\ell = 4$ and $\ell = 5$, are presented in table IV together with the associated ratios m/η. The modal value for $\ell = 4$ is 2/3, and the modal value for $\ell = 5$ is 1/2. These values suggest the following total procedure: Perform the chain pooling ANOVA with $m = 1$ and the preferred α_p for the guessed value of η. The result will be an estimate $\hat{\eta}$ of η. If $\ell = 4$, let m be the integer nearest $(2/3)\hat{\eta}$. If $\ell = 5$, let m be the integer nearest $(1/2)\hat{\eta}$. With this new value of m, perform a second ANOVA of the observations with $\alpha_p = 1.0$. In the case of $\ell = 6$, the ANOVA with $m > 1$ would be performed with $\alpha_p = 1.0$, $\alpha_f = 0.001$, and a value of m based on $\hat{\eta}$, where the exact value of m is chosen to control $\overline{R}_1(\eta)$.

OPERATING CHARACTERISTICS

The conservative strategy of using $m = 1$ has been shown to be a strategy that might well be used in a preliminary chain pooling ANOVA. The ANOVA with $m = 1$ would then give an estimate $\hat{\eta}$ of η that leads to a larger value of m, perhaps $m \cong (2/3)\hat{\eta}$. The main interest is in the operating characteristics for $m > 1$. Risk curves for preferred values of m and α_p are illustrated in figures 8(a), 10, and 11(a) and (b). These curves map out detailed values of $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$ for a wide variety of values of η and λ, but they do not give a clear picture of the response of $\overline{R}_2(\lambda)$ to η and λ. The response of $\overline{R}_2(\lambda)$ to η and λ is shown by plots of $\overline{R}_2(\lambda)$ against λ for stated values of $\overline{R}_1(\eta)$ and η. The plots were obtained by reading curves

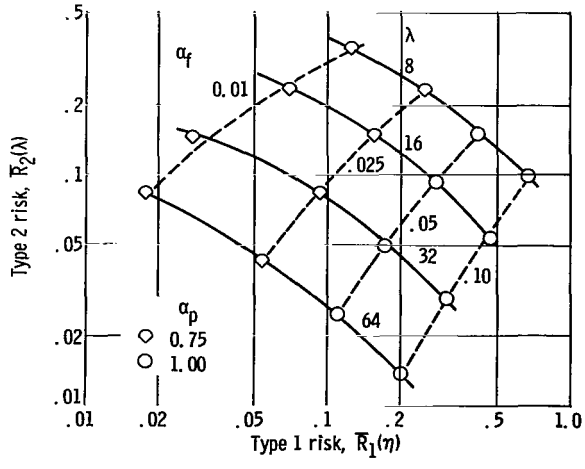
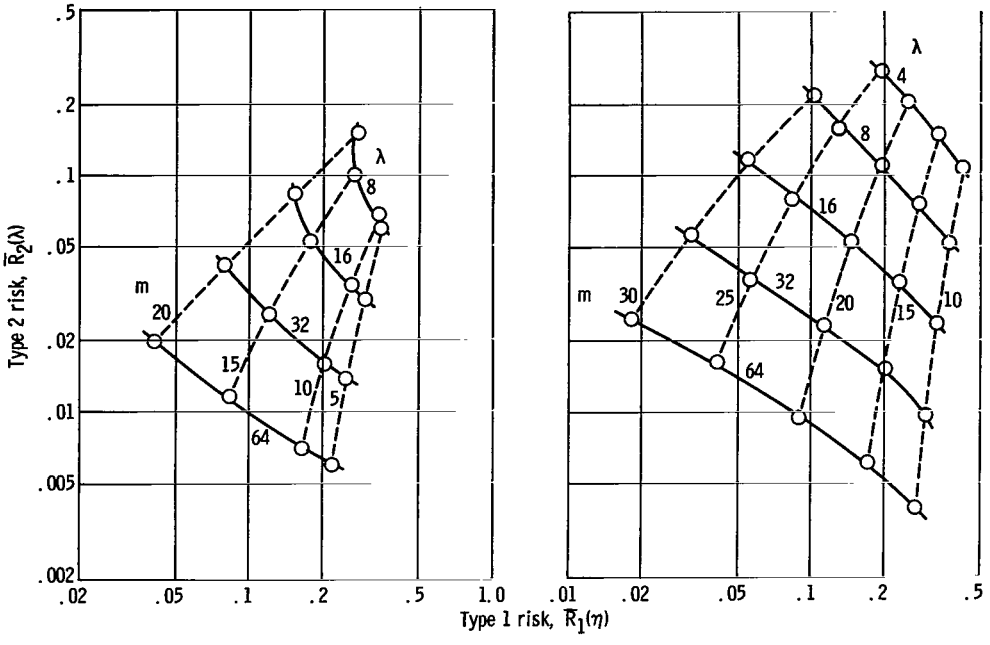


Figure 10. - Risk curves. $\ell = 4$, $\rho = 11$, $\eta = 4$, and $m = 3$.



(a) $\rho = 40$, $\eta = 23$.

(b) $\rho = 30$, $\eta = 33$.

Figure 11. - Risk curves. $\ell = 6$, $\alpha_p = 1.00$, and $\alpha_f = 0.001$.

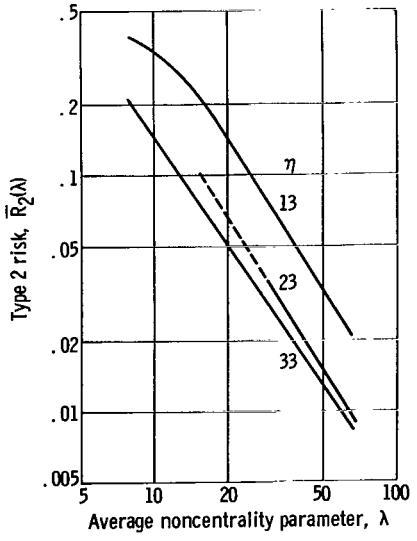


Figure 12. - Type 2 risks. $\ell = 6$, $\alpha_p = 1.0$, $\alpha_f = 0.001$, and $\bar{R}_1(\eta) = 0.10$.

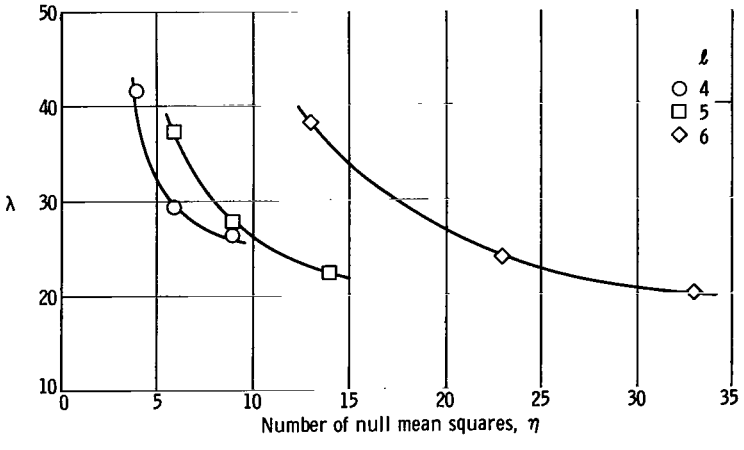


Figure 13. - Operating characteristics with preferred m. $\bar{R}_1(\eta) = 0.10$ and $\bar{R}_2(\lambda) = 0.05$.

such as those in figures 8(a), 10, and 11(a) and (b) for the given values of η at particular abscissa values; namely at $\overline{R}_1(\eta) = 0.05$, 0.10, and 0.20. The values of $\overline{R}_2(\lambda)$ so read were plotted as functions of λ, as illustrated in figure 12. This plot shows that for preferred m and for fixed $\overline{R}_1(\eta)$, the values of $\overline{R}_2(\lambda)$ are strongly dependent on λ and are only weakly dependent on η. This observation is the basis for using the symbol $\overline{R}_2(\lambda)$ to express the fact that $\overline{R}_2$ is always a function of λ but is mostly independent of η.

The general performance of the preferred strategies ($m > 1$) was illustrated by figure 12. This plot shows type 2 risks as a function of the average noncentrality parameter λ for stated levels of ℓ, η, and the type 1 risks. These curves permit reading values of λ for specified values of the type 2 risks. Values of λ for values of $\overline{R}_2(\lambda)$ of 0.05, 0.10, and 0.20 were read and plotted as functions of η for stated values of $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$, as illustrated by figure 13. For preferred m, figure 13 shows that, at the stated risks $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$, the detectable values of λ generally decrease with increasing η.

EFFICIENCY AS FUNCTION OF EXPERIMENT SIZE

The purpose of the multiple decision procedure with respect to a 2^ℓ design experiment is to detect the nonnull populations ρ in number. The quantity signaling that a population is nonnull is the square of the effect parameter, μ_i^2; and the relative strength, or signal to noise ratio is μ_i^2/σ^2. From equation (5), this ratio is also given by $\theta^2\delta_i^2$. Because of equation (6), the average value of μ_i^2/σ^2 was θ^2. At given levels of risks, $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$, the detection of the experiment is defined as the number of signals to be detected ρ divided by the ratio of the average relative signal to noise θ^2. The detection efficiency ψ is defined as the detection divided by the number of observations

$$\psi = \frac{\rho}{\theta^2 2^\ell}$$

and from equation (12),

$$\psi = \frac{\rho}{\lambda} \tag{15}$$

The quantity ψ thus provides a measure for comparing the efficiency of 2^4, 2^5, and 2^6 fractional factorial experiments with each other. The quantity λ for given risk levels was plotted as curves, such as those of figure 13. The associated values of η

have equivalent values of ρ according to $\rho = 2^{\ell} - \eta - 1$. Values of λ read from such curves, together with the associated values of ρ, were substituted in equation (15) to produce the values of ψ, the detection efficiency.

This efficiency was shown to increase very rapidly from 2^4 to 2^6 design experiments. One implication of this result is that the experimenter should include all conceivable variables in his first design with no intention of adding other variables to the investigation at a later date. Such a policy may result in a large number of test conditions, leading the experimenter to seek means of achieving economy. One obvious method of achieving economy is to avoid the use of replication, in which case the methods of chain pooling become essential for analyzing results. Another method of achieving economy is to use fractional replication, but fractional replication creates questions as to whether significant interactions are being excluded from the model. Furthermore, large designs may not be performable under homogeneous conditions, and blocked designs may be needed. If a severely fractionated design is performed, and the experimenter wishes to augment the testing to evaluate additional interactions, special relations must exist between the old and the new test conditions so that the newly evaluated interactions will not be confounded with block effects. Sequences of designs that satisfy these relations are given in reference 5. The sequences are such that observations from the first block can be used to estimate the coefficients of a simple model and then be retained and combined with observations from new blocks so that all acquired observations are used cumulatively to estimate models of successively greater generality.

ROBUSTNESS

The subject of robustness was investigated by generating variates without transformation from the rectangular to the approximate normal distribution. The operating characteristics of the chain pooling procedure for the pseudorectangular distribution were in good agreement (ref. 9) with those for the pseudonormal distributions.

CHARTS FOR CONTROL OF TYPE 1 RISKS

Classical hypothesis testing enables the statistician to test an alternative hypothesis against a null hypothesis with a predetermined bound on the probability of a type 1 error. This natural formulation of a decision procedure can sometimes be applied to multiple decision procedures. In the case of chain pooling, the analogous procedure of putting a predetermined bound on the type 1 risks is not completely feasible. Partial feasibility is illustrated by figure 14. The choice, for example, of a strategy (m, α_p, α_f) leads

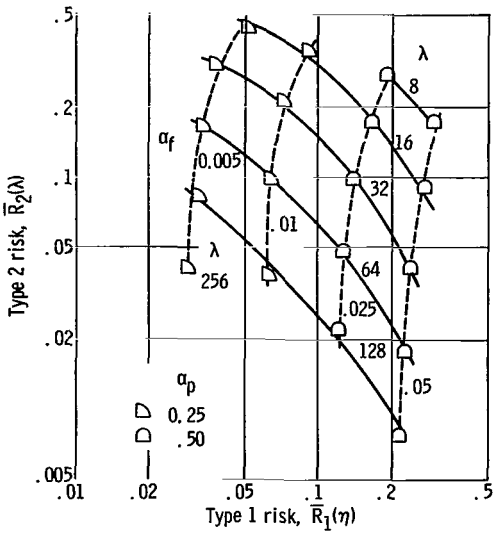


Figure 14. - Risk curves. $\ell = 4$, $\rho = 9$, $\eta = 6$, and $m = 1$.

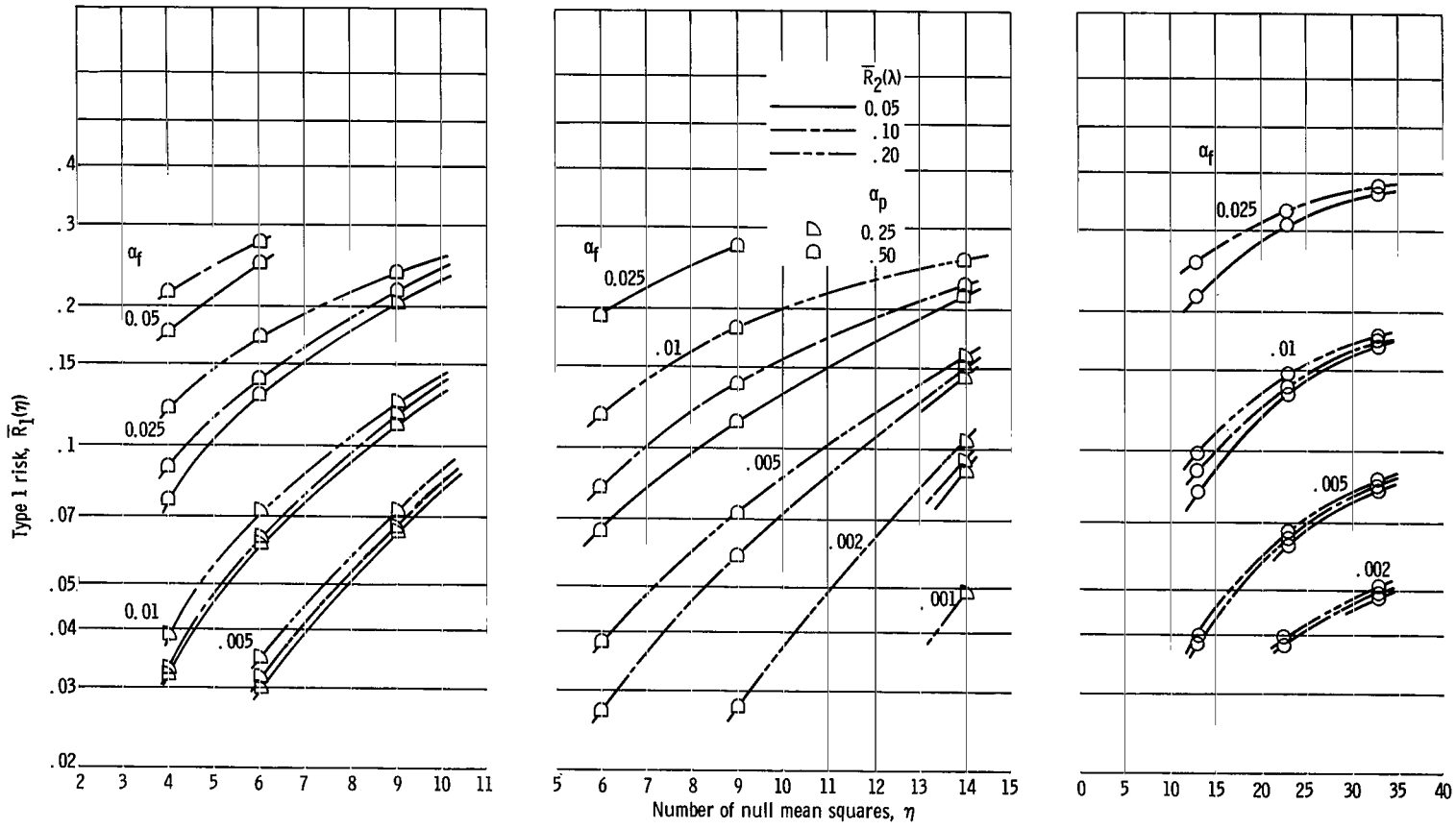
to one of the dashed curves. These curves are seen to be nearly vertical for $\lambda \geq 64$. Thus, the choice of a fixed strategy, for instance, $(m, \alpha_p, \alpha_f) = (1, 0.25, 0.01)$, would essentially fix $\bar{R}_1(\eta)$ at 0.06 (fig. 14) provided that $\lambda > 64$; however, for $\lambda < 64$, $\bar{R}_1(\eta)$ would increase with decreasing λ.

Figure 14 illustrates a situation where the type 1 losses are insensitive to λ, if λ is sufficiently large. The type 2 risks had previously been shown to be mainly a function of λ. Figure 14 thus illustrates a situation where the type 1 risks are independent of the type 2 risks, if the type 2 risks are sufficiently small. This fact is further illustrated by figure 15(a). This figure was obtained by reading the risk curves of reference 9, as illustrated by figure 14, to obtain values of $\bar{R}_1(\eta)$ that correspond to the specific type 2 risk levels of $\bar{R}_2(\lambda) = 0.05$, 0.10, and 0.20. The result is that the three curves for any given α_f lie fairly close to one another independently of the three levels of $\bar{R}_2(\lambda)$. Figure 15(a) thereby provides the information by which an a priori guess of η can be used to select a value of α_f that should control the level of $\bar{R}_1(\eta)$ fairly independently of λ.

Risk curves with $m = 1$ and $\ell = 5$ are given in reference 9. These curves were read at values of $\bar{R}_2(\lambda)$ of 0.05, 0.10, and 0.20 to produce the curves of figure 15(b). This figure also suggests that, with an a priori choice of η, the value of $\bar{R}_1(\eta)$ would not be especially sensitive to $\bar{R}_2(\lambda)$ or to λ.

A similar reading of curves for $\ell = 6$, as given in reference 9, produced the curves of figure 15(c). The relation between $\bar{R}_1(\eta)$ and η is again seen to be essentially independent of $\bar{R}_2(\lambda)$.

With $m > 1$, the control of $\bar{R}_1(\eta)$ is much more difficult. For example, the lines

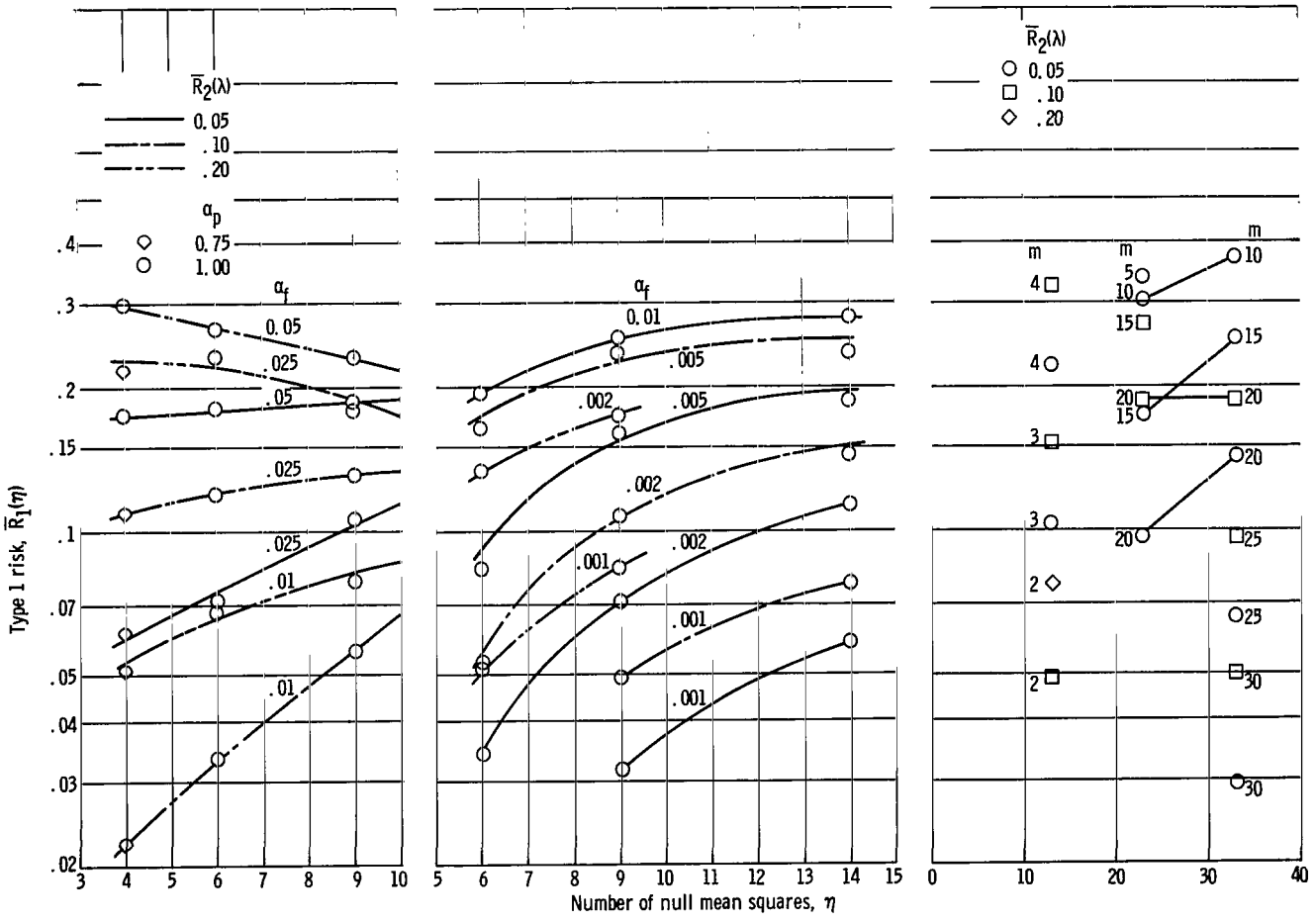


(a) $\ell = 4$.

(b) $\ell = 5$.

(c) $\ell = 6$, $\alpha_p = 1$.

Figure 15. - Curves for control of type 1 risks with $m = 1$ and preferred α_p.



(a) $\ell = 4$.

(b) $\ell = 5$, $\alpha_p = 1.0$.

(c) $\ell = 6$, $\alpha_p = 1.0$, $\alpha_f = 0.001$.

Figure 16. - Type I risks with preferred m and α_p.

of the constant α_f in figure 10 have no regions where a fixed strategy results in essentially constant $\overline{R}_1(\eta)$. The same technique used in plotting figure 15 was used to plot figure 16 from data in reference 9 for strategies where both m and α_p were at preferred values. Here the values of $\overline{R}_1(\eta)$ are widely separated according to the value of $\overline{R}_2(\lambda)$.

A rigid relation between the type 1 and type 2 risks was illustrated by figure 14 in the region of $\lambda > 64$; that is, the choice of m, α_p, and α_f essentially fixed the value of $\overline{R}_1(\eta)$ independently of $\overline{R}_2(\lambda)$, as shown by the vertical direction of the curves of the constant α_f. An elastic relation between the type 1 and type 2 risks is illustrated in figure 10, in that for any fixed strategy (m, α_p, α_f), $\overline{R}_1(\eta)$ varies with $\overline{R}_2(\lambda)$ in any range of λ. An increase in the value of λ will simultaneously reduce $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$. In this situation, a statistician might desire some bound on the average probability of failing to detect real effects. Assume that he has estimated η from an initial ANOVA with m = 1. He can now enter figure 16(a) with this estimate, $\hat{\eta}$, and with the desired type 2 loss bound (particular value of $\overline{R}_2(\lambda)$). For such values of $\hat{\eta}$ and $\overline{R}_2(\lambda)$, he can now choose a value of α_f to give some desired $\overline{R}_1(\eta)$. He can then say that if the average noncentrality parameter is large enough to hold the type 2 risks to the chosen bound, the selected α_f will be small enough to hold the type 1 risks to the chosen $\overline{R}_1(\eta)$. If the average noncentrality parameter is larger than what he had hoped for, both the type 2 and type 1 risks will be lower than specified. If the average noncentrality parameter is less than what would give small type 2 risks, he must accept increased type 1 risks. In any event, having chosen some strategy ($m > 1$, α_p, α_f), he must realize that it is an optimal strategy for some combination of the type 1 and type 2 risks, the only drawback being that, if the average noncentrality parameter is small enough to boost the type 2 risks, it will also boost the type 1 risks.

This elasticity between the type 1 and type 2 risks with $m > 1$ could be considered an advantage over the strategy with m = 1; that is, if the average noncentrality parameter is smaller than what the statistician had hoped for, then, with the elastic type 1 risk, the type 2 risks will be smaller than they would have been if the type 1 risk had been rigidly controlled by the strategy of m = 1. (Compare figs. 10 and 14.)

CONCLUDING REMARKS

Factorial experiments are essential when interactions among the factors can be important. Such experiments might be performed without replication, if the experiments are expensive, as in alloy development, or in the destructive testing of structures, or where many variables are involved, such as high-temperature protective coating

research. Not replicating produces an economy but leads to a situation where there is no well-established strategy for analyzing the results.

Some Monte Carlo investigations of varieties of chain pooling procedures for 2^4, 2^5, and 2^6 fractional factorial design experiments suggested that a good overall strategy consists of starting the analysis with just the single smallest mean square, and it is assumed to have come from a null population. Testing proceeds at some level α_p in the order of increasing magnitude of the mean squares. Mean squares found not significant are pooled into the denominator of the test statistic. If a mean square is found significant at level α_p, testing begins with it at a more stringent level α_f and proceeds in the order of increasing magnitude of the mean squares. If a mean square is found significant at level α_f, it and all larger mean square are declared significant. This procedure gives an estimate $\hat{\eta}$ of the number of null effects in the experiment. A second analysis of the data is then performed with the use of a test statistic with $m > 1$ smallest mean squares initially pooled into the denominator of the test statistic.

For 2^4 design experiments, a good choice of m is the integer nearest $(2/3)\hat{\eta}$. For 2^5 design experiments a good choice of m is the integer nearest $(1/2)\hat{\eta}$. For 2^6 design experiments, a good strategy consists of choosing $\alpha_p = 1.0$, $\alpha_f = 0.001$, and then choosing m to control the type 1 risk.

Curves are given so that some control over the average probability of type 1 error can be achieved through the choice of α_f.

The actual type 1 and type 2 error probabilities depend on the magnitudes of the real effects, and curves are given for estimating weighted average error probabilities after the real effects have been estimated.

The chain pooling procedure compared favorably with Daniel's modulus ratio statistic (half-normal plotting) for the case of just one real effect.

Lewis Research Center,
National Aeronautics and Space Administration,
Cleveland, Ohio, July 6, 1967,
129-03-01-03-22.

APPENDIX A

CURVES FOR POSTERIOR ESTIMATION OF RISKS

The assumptions are now introduced that a 2^{ℓ} design experiment has been completed, that a preliminary chain pooling ANOVA has been performed with $m = 1$ and with some α_p and α_f, and that a second ANOVA has been performed with some preferred $m > 1$. The experimenter may be satisfied with the results, or he may have some questions such as

(1) Should the ANOVA be performed at some other α_f?

(2) Should the conditions of the experiment be changed to provide larger values of the μ_i?

(3) Should more precise (and more costly) instrumentation be used?

Partial answers to these questions may be obtained from posterior estimates of the type 1 and type 2 risks. The starting point of such estimates is the estimation of an average noncentrality parameter λ.

An average of the λ_i, as computed from equation (3), is

$$\lambda = \frac{1}{\rho} \sum_{i=1}^{\rho} \lambda_i$$

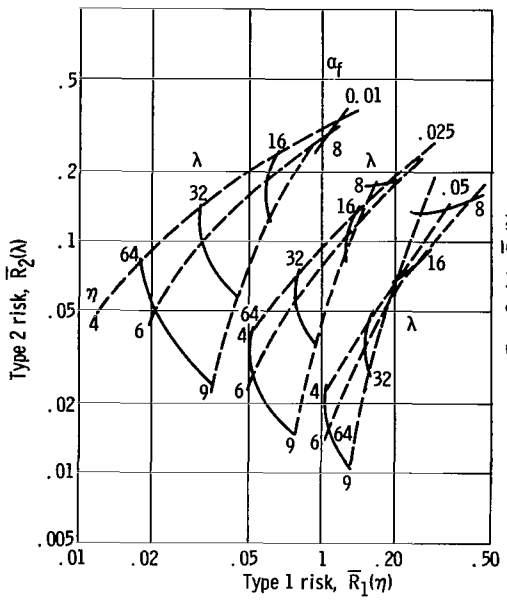
$$= \frac{1}{\rho} \sum_{i=1}^{\rho} 2^{\ell} \left(\frac{\mu_i}{\sigma} \right)^2$$

$$= \frac{2^{\ell}}{\rho \sigma^2} \sum_{i=1}^{\rho} \mu_i^2$$

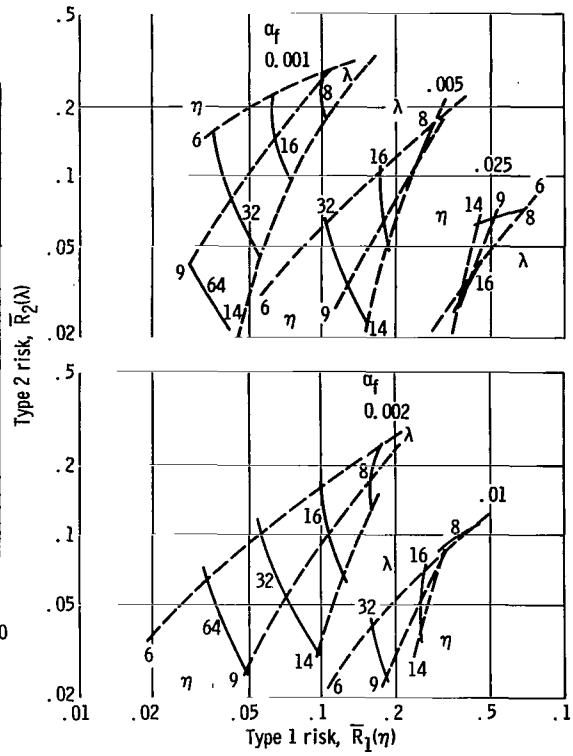
Replacing the unknown parameters by quantities estimated from the experiment gives

$$\hat{\lambda} = \frac{2^{\ell}}{\hat{\rho} \hat{\sigma}^2} \sum_{j=1+\hat{\eta}}^{2^{\ell}-1} \hat{\mu}_j^2 \tag{A1}$$

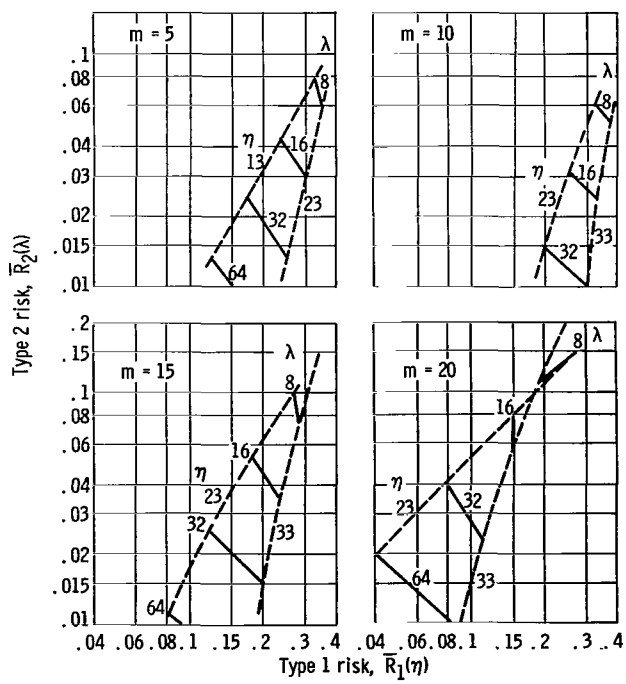
The multiple decision ANOVA procedure leads to the conclusion that $\hat{\eta}$ mean squares



(a) $\ell = 4$, preferred m and α_p.



(b) $\ell = 5$, preferred m and α_p.



(c) $\ell = 6$, $\alpha_p = 1.00$, and $\alpha_f = 0.001$.

Figure 17. - Posterior risk curves.

are free from real effects, and that $\hat{\rho} = 2^{\ell} - \hat{\eta} - 1$ mean squares contain real effects. The $\hat{\rho}$ effects concluded to be significant have estimators $\hat{\mu}_i$ give by Yates' algorithm.

The estimate $\hat{\sigma}^2$ can be obtained from the procedure given by Wilk, Gnanadesikan and Freeny (ref. 3). Where $\hat{\eta}$ is regarded as the total sample size, the censored sample consisting of the m smallest ordered contrasts is used to estimate σ^2.

Information such as that presented in figures 8(a), 10, and 11(a) and (b) is presented in a different form in figure 17. The statistician may enter figure 17 with (1) the value of ℓ for the design of the experiment, (2) the value of α_f used in the decision making, (3) the value of $\hat{\eta}$ resulting from the decision procedure, and (4) the average value $\hat{\lambda}$ computed as just described. Visual interpolation with respect to η and λ should then result in a pair of values $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$ that would be the posterior risk estimates for the total procedure. In the event that the values of $\hat{\eta}$ and $\hat{\lambda}$ did not fall within the ranges of existing curves in figure 17, the statistician could still notice whether his values of $\hat{\eta}$ and $\hat{\lambda}$ suggested that the risks were less than, or greater than, some values for which curves had been drawn. Figure 17 was obtained from Monte Carlo computations where the true values of η and λ were known. The statistician's posterior estimates of $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$ will be dependent on errors of estimate in $\hat{\eta}$ and $\hat{\lambda}$. Furthermore, if the distribution of λ_i in the real experiment was favorable to the decision procedure, the estimate of $\overline{R}_2(\lambda)$ will be too high because figure 17 was based on unfavorable distributions of the λ_i.

The value to the experimenter of making these posterior risk estimates is twofold: (1) For the strategy with $m > 1$, the type 1 risk is dependent on the value of λ so that the selection of a strategy (m, α_p, α_f) might result in a true $\overline{R}_1(\eta)$ somewhat different from what the statistician had desired when he first selected a strategy according to figure 16. The posterior estimation of $\overline{R}_1(\eta)$ might then tell the statistician to perform another analysis with a different value of α_f. (2) The analysis with $m > 1$ is done with no knowledge of λ and therefore with no attempt to balance the type 1 and type 2 risks. In some cases, the experimenter may want to do some balancing. Suppose the value of λ were small so that the posterior risk estimates were $\overline{R}_1(\eta) = 0.20$ and $\overline{R}_2(\lambda) = 0.20$. Then, further analysis would seem to be of no value. On the other hand, if $\hat{\lambda}$ were large and the posterior risk estimates were, for example, $\overline{R}_1(\eta) = 0.20$ and $\overline{R}_2(\lambda) = 0.02$, the experimenter might desire another analysis (presumably with decreased α_f or increased m) that would decrease $\overline{R}_1(\eta)$ at the expense of increasing $\overline{R}_2(\lambda)$.

APPENDIX B

ILLUSTRATIVE EXAMPLE - COBALT-BASE-ALLOY DEVELOPMENT, ONE-HALF REPLICATE WITH FIVE FACTORS

General Chain Pooling ANOVA Strategies

The analysis is presented from two points of view. The first point of view, the significance point of view, assumes that the statistician must be conservative in drawing conclusions, and that the probability of making type 1 errors will be controlled. The second point of view, the screening point of view, assumes that all effects that might be real should be so identified. With this point of view, the statistician is particularly concerned about type 2 errors.

In either case, the ANOVA proceeds in two analyses. The first analysis begins with the strategy $m = 1$ and leads to an estimate $\hat{\eta}$ of the number of null mean squares. Based on this estimate of η, a second iteration of the ANOVA is performed with preferred $m > 1$.

The results of the Yates' computation are given in table V. The responses y_i are logarithms of stress-rupture times to failure of cast specimens. The effects to be estimated (table V) include 1 grand mean, 5 main effects, and 10 interactions.

The strategy (m, α_p, α_f) is chosen according to information on the operating characteristics given for $m = 1$ by figure 15(a) and for preferred $m > 1$ by figure 16(a). The use of these curves requires that a prior estimate be made for η. For this example, the number of null mean squares is assumed to be equal to the number of highest order interactions; namely, the assumption is $\eta = 10$.

In addition to making a prior estimate of η, the use of the curves of figures 15 and 16 also requires that some assumptions be made about the value of $R_2(\lambda)$. Because the operating characteristics with $m = 1$ are not sensitive to the value of λ, the curves of figure 15 are closely spaced with respect to $R_2(\lambda)$, and therefore a single arbitrary value (the middle curve for $\overline{R}_2(\lambda) = 0.10$) will be used. The operating characteristics are critically dependent on λ or $\overline{R}_2(\lambda)$ when $m > 1$, as in figure 16. In using figure 16, assumptions are made about $\overline{R}_2(\lambda)$ according to the purpose of the ANOVA. Comparatively large type 2 risks should be anticipated with the significance point of view and for such a situation, the assumption is $\overline{R}_2(\lambda) = 0.20$. With the screening point of view, comparatively small type 2 risks should be anticipated and $\overline{R}_2(\lambda) = 0.05$ is assumed. These assumptions of risk levels of 0.05, 0.10, or 0.20 are higher than the significance levels often used in statistical procedures. These high type 2 risk levels are regarded as being consistent with the election to use the economy model experiment

TABLE V. - ILLUSTRATIVE EXAMPLE OF ONE-HALF 2^5 DESIGN EXPERIMENT

i, j	Treatment levels	Responses, y_i	Parameters	Parameter estimates,[a] $\hat{\mu}_i$	Mean squares, Z_j	Significance point of view					
						First analysis with (m, α_p, α_f) = 1, 0.25, 0.01			Second analysis with (m, α_p, α_f) = 5, 1.0, 0.01		
						U_j(exper.)	$U_j(\alpha_p)$	$U_j(\alpha_f)$	U_j(exper.)	$U_j(\alpha_p)$	$U_j(\alpha_f)$
0	(1)	2.2715	μ_0	1.6522							
1	ae	2.0708	μ_1	-.0297	0.000079						
2	be	1.3745	μ_2	-.3833	.000103	1.1319	1.923	1.99986			
3	ab	1.2458	μ_{12}	.0781	.004408	[b]2.8810	2.527	2.9809			
4	ce	2.2810	μ_3	.1002	.007567	2.9295					
5	ac	2.0950	μ_{13}	-.0166	.014160	2.9619					
6	bc	1.5207	μ_{23}	-.0025	.018099	2.9701			2.4449		5.31
7	abce	1.5290	$-\mu_{45}$	-.0217	.020248	2.9733			2.6090		
8	de	1.8199	μ_4	-.1464	.021949	2.9753			2.7285		
9	ad	1.5687	μ_{14}	.0336	.028594	[c]2.9810			3.1244		
10	bd	.7947	μ_{24}	-.0022	.032091				3.2966		
11	abde	1.2701	$-\mu_{35}$	.0448	.043254				3.7303		
12	cd	2.2006	μ_{34}	.0423	.097554				4.7253		
13	acde	1.9759	$-\mu_{25}$	-.0356	.160510				5.1548		
14	bcde	1.1924	$-\mu_{15}$	-.0520	.342750				[c]5.5722		
15	abcd	1.2240	$-\mu_5$	-.0370	2.350200						

[a] $\hat{\mu}_i$ are contrasts divided by 2^{ℓ}.
[b] Significant at level α_p.
[c] Significant at level α_f.

that has no replication. Of course, the parameters of the true physical situation could be such that these assumptions are much too large or much too small. Such assumptions must be made so that a decision strategy can be selected that will be somewhat appropriate to the experimenter's needs in the design and analysis of the experiment. When the analysis is completed, posterior estimates can be made of the risks $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$.

Because the type 2 risks should always be minimized, the largest α_f value that will not result in too large a value of $\overline{R}_1(\eta)$ should be chosen. Since the operating characteristics with $m = 1$ are inferior to those with preferred $m > 1$, the statistician should be less stringent about $\overline{R}_1(\eta)$ when using $m = 1$, than when using preferred $m > 1$. Under the previously stated assumptions concerning $\overline{R}_2(\lambda)$, the strategy (m, α_p, α_f) will be chosen from figure 15(a) and from 16(a) according to the resulting value of $\overline{R}_1(\eta)$.

Significance Point of View

First analysis, $m = 1$:

(1) Assume that $\eta = 10$ and $\overline{R}_2(\lambda) = 0.10$.

(2) In figure 15(a) with $\eta = 10$ and $\overline{R}_2(\lambda) = 0.10$, choose the highest α_f that yields an acceptable $\overline{R}_1(\eta)$ and note the required α_p. Figure 15(a) shows that if $\alpha_f = 0.01$, $\overline{R}_1(\eta)$ is 0.14 and α_p is required to be 0.25.

(3) The strategy $(m, \alpha_p, \alpha_f) = (1, 0.25, 0.01)$ and the chain pooling ANOVA resulted in $\hat{\eta} = 8$ and $\hat{\rho} = 7$.

Second analysis, $m \cong (2/3)\hat{\eta} = (2/3)8 \cong 5$:

(1) Assume that $\eta = \hat{\eta} = 8$ and that $\overline{R}_2(\lambda) = 0.20$.

(2) In figure 16(a) with $\eta = 8$ and $\overline{R}_2(\lambda) = 0.20$, choose the highest α_f that yields an acceptable $\overline{R}_1(\eta)$ and note the required α_p. Figure 16(a) shows that if $\alpha_f = 0.01$, the $\overline{R}_1(\eta)$ is 0.08, and α_p is required to be 1.00.

(3) The strategy $(m, \alpha_p, \alpha_f) = (5, 1.00, 0.01)$, and the chain pooling ANOVA resulted in $\hat{\eta} = 13$ and $\hat{\rho} = 2$.

Screening Point of View

First analysis, $m = 1$:

(1) Assume that $\eta = 10$ and that $\overline{R}_2(\lambda) = 0.10$.

(2) In figure 15(a) with $\eta = 10$ and $\overline{R}_2(\lambda) = 0.10$, choose the highest α_f that yields an acceptable $\overline{R}_1(\eta)$ and note the required α_p. Figure 15(a) shows that if $\alpha_f = 0.025$, the $\overline{R}_1(\eta)$ is 0.24 and α_p is required to be 0.50.

(3) The strategy $(m, \alpha_p, \alpha_f) = (1, 0.50, 0.025)$, and the chain pooling ANOVA resulted in $\hat{\eta} = 4$ and $\hat{\rho} = 11$.

Second analysis, $m \cong (2/3)\hat{\eta} = (2/3)4 \cong 3$:

(1) Assume that $\eta = 4$ and that $\overline{R}_2(\lambda) = 0.05$.

(2) In figure 16(a) with $\eta = 4$ and $\overline{R}_2(\lambda) = 0.05$, choose the highest α_f that yields an acceptable $\overline{R}_1(\eta)$ and note the required α_p. Figure 16(a) shows that if $\alpha_f = 0.05$, the $\overline{R}_1(\eta)$ is 0.17 and α_p is required to be 1.00.

(3) The strategy $(m, \alpha_p, \alpha_f) = (3, 1.00, 0.05)$, and the chain pooling ANOVA resulted in $\hat{\eta} = 11$ and $\hat{\rho} = 4$.

In summary, the significance point of view with $m = 1$ resulted in $\hat{\rho} = 7$, whereas the preferred $m = 5$ resulted in $\hat{\rho} = 2$. The screening point of view with $m = 1$ resulted in $\hat{\rho} = 11$, whereas the preferred $m = 3$ resulted in $\hat{\rho} = 4$. The need for going beyond $m = 1$ to the second iteration with $m > 1$ is obvious. The statistician might seek convergence of $\hat{\eta}$ by continuing with further iterations of the ANOVA, using new

values of $m \cong (2/3)\hat{\eta}$, where $\hat{\eta}$ is determined by the immediately preceding iteration. The operating characteristics of such a procedure have not been investigated.

Posterior Estimates of Risks

The risk levels of figure 16 for preferred $m > 1$ are dependent on the value of η, which was only crudely estimated by the initial analysis with $m = 1$. Furthermore, the risk levels of figure 16 are dependent on the values of the λ_i, and no information on their values was used in entering figure 16. For these two reasons, a posterior estimate of $\overline{R}_1(\eta)$ and $\overline{R}_2(\lambda)$ is highly desirable. Such estimates will now be made for the second iteration under the screening point of view.

The quantity σ^2 is estimated in the manner of reference 3, which requires a decision to use some small number of the smallest mean squares. The value of $m = 3$ as used in the ANOVA is retained for this estimate of σ^2. The equivalence between the notation of reference 3 and the present notation is

$$\frac{K}{M} = \frac{\hat{\eta}}{m} = \frac{11}{3} = 3.67$$

Also,

$$S_M^2 = \sum_{j=1}^{m} \frac{Z_j}{mZ_m}$$

From table V, the sum of the three smallest mean squares is 0.004590. Also,

$$mZ_m = (3)\,(0.004408) = 0.013224$$

$$S_M^2 = \frac{0.004590}{0.013224} = 0.347$$

Interpolation in table 1 of reference 3 gives $1/\hat{\xi} = 8.30$, and the variance estimate is

$$\hat{\sigma}^2 = \frac{Z_m}{\hat{\xi}} = (0.004408)\,(8.30) = 0.0366$$

TABLE VI. - ONE-HALF 2^5 DESIGN EXPERIMENT ON COBALT ALLOYS

Analysis	Prior guess of number of null mean squares η	Prior type 1 risk, $\bar{R}_1(\eta)$	Assumed type 2 risk, $\bar{R}_2(\lambda)$	Number of mean squares pooled before testing, m	Nominal size of preliminary pooling test, α_p	Nominal size of final significance test, α_f	Estimate of number of null mean squares, $\hat{\eta}$	Estimate of number of real effects, $\hat{\rho}$	Revised m	Estimate of average noncentrality parameter, $\hat{\lambda}$	Posterior type 1 risk, $\bar{R}_1(\eta)$	Posterior type 2 risk, $\bar{R}_2(\lambda)$
Significance point of view												
First	10	0.14	0.10	1	0.25	0.01	8	7	5	----	-----	-----
Second	8	.08	.20	5	1.00	.01	13	2	---	23.6	[a]0.07	[a]0.07
Screening point of view												
First	10	0.24	0.10	1	0.50	0.025	4	11	3	----	-----	-----
Second	4	.17	.05	3	1.00	.05	11	4	---	20.1	0.18	0.06

[a]Minor extrapolation.

The ANOVA results were $\hat{\eta} = 11$ and $\hat{\rho} = 4$. The significant effects are therefore the four largest absolute values of the $\hat{\mu}_i$ (aside from $\hat{\mu}_0$) of table V

Estimate	Value
$\hat{\mu}_2$	-0.3833
$\hat{\mu}_4$	-.1464
$\hat{\mu}_3$	.1002
$\hat{\mu}_{12}$	.0781

The mean noncentrality parameter is given by equation (A1):

$$\hat{\lambda} = \frac{2^{\ell}}{\hat{\rho}\hat{\sigma}^2} \sum_{j=1+\hat{\eta}}^{2^{\ell}-1} \hat{\mu}_j^2$$

$$= \frac{16}{(4)\,(0.0366)} \left[(0.3833)^2 + (0.1464)^2 + (0.1002)^2 + (0.0781)^2\right]$$

$$= 20.164$$

The posterior risk estimates are obtained from figure 17(a). The strategy used $\alpha_f = 0.05$ to conclude that $\hat{\eta} = 11$. From figure 17(a) with $\lambda = 20$, $\eta = 11$, and $\alpha_f = 0.05$, crude graphical extrapolation suggests that, roughly, $\overline{R}_1(\eta) = 0.18$ and $\overline{R}_2(\lambda) = 0.06$.

The four analyses of the 2^4 experiment on cobalt base-alloys are summarized in table VI. In review, the analysis was begun from the significance point of view with $m = 1$ and with the guess of $\eta = 10$. Because the type 1 errors are insensitive to λ with $m = 1$, an arbitrary value could be assumed for $\overline{R}_2(\lambda)$, and the value chosen was $\overline{R}_2(\lambda) = 0.10$. Under these conditions, the choice of $(m, \alpha_p, \alpha_f) = (1, 0.25, 0.01)$ and the use of figure 15(a) gives the estimate $\overline{R}_1(\eta) = 0.14$. Performance of the analysis resulted in $\hat{\eta} = 8$ and $\hat{\rho} = 7$, and for this 2^4 experiment, a preferred value of m was then chosen according to table V as $m \cong (2/3)\hat{\eta} = (2/3)(8) \cong 5$. The second analysis with $m = 5$ was begun with a prior estimate of η from the first analysis, namely, $\hat{\eta} = 8$. From a significance point of view, the type 2 error should be expected to be high, and (consistent with this point of view) the type 2 risk was arbitrarily assumed to be large, namely, $\overline{R}_2(\lambda) = 0.20$. The decision procedure with $(m, \alpha_p, \alpha_f) = (5, 1.00, 0.01)$ then begins with $\overline{R}_1(\eta)$ estimated from figure 16(a) as $\overline{R}_1(\eta)$ estimated from figure 16(a) as $\overline{R}_1(\eta) = 0.08$. The procedure resulted in $\hat{\eta} = 13$ and $\hat{\rho} = 2$. The mean squares selected as being significant (aside from the grand mean) are the $\hat{\rho} = 2$ largest mean squares. The posterior estimation of risks using figure 17(a) then leads to $\overline{R}_1(\eta) = 0.07$ and $\overline{R}_2(\lambda) = 0.07$. Thus, the posterior estimate of $\overline{R}_1(\eta)$ is fairly close to the initial control value of $\overline{R}_1(\eta) = 0.08$; however, the posterior estimate of $\overline{R}_2(\lambda) = 0.07$ (which has now been based on $\hat{\rho}$ and $\hat{\lambda}$) is quite different from the arbitrary initial assumption of $\overline{R}_2(\lambda) = 0.20$ (which was chosen with no knowledge of λ).

As presented in table VI, the posterior risk estimates under the significance point of view were based on $\hat{\eta} = 13$ and $\hat{\rho} = 2$, whereas the screening point of view resulted in $\hat{\eta} = 11$ and $\hat{\rho} = 4$. Thus, the strategy used with the screening point of view has doubled the number of effects concluded to be significant.

The posterior risks for the screening point of view, when compared with the significance point of view, showed that increasing $\overline{R}_1(\eta)$ from 0.07 to 0.18 achieved a reduction of $\overline{R}_2(\lambda)$ from 0.07 to 0.06. That such a small improvement in $\overline{R}_2(\lambda)$ occurred at such a large cost of $\overline{R}_1(\eta)$ might be associated with the fact that $\overline{R}_2(\lambda)$ was already at a point of diminishing returns in the analysis with the significance point of view. Of course, the important conclusion with respect to the experiment is that at least $\hat{\rho} = 2$ effects (aside from the grand mean) are clearly significant, whereas as many as $\hat{\rho} = 4$ effects are possibly significant. The reader may compare these conclusions with his interpretation of the half-normal plot, which is shown in figure 18.

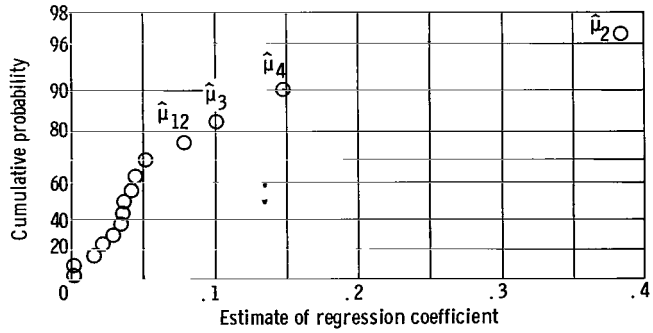


Figure 18. - Half-normal plot for one-half 2^5 design experiment on cobalt-base alloys.

APPENDIX C

ILLUSTRATIVE EXAMPLE - PENICILLIN PRODUCTION, FIVE-FACTOR EXPERIMENT IN TWO BLOCKS

This experiment was first analyzed in reference 1. It was also used as an example of half-normal plotting in reference 2 and as an example of variance estimation in reference 3. The results of a chain pooling ANOVA together with posterior estimates of type 1 and type 2 risks are given in table VII. These results will be compared, for equal levels of the type 1 risk, with results achieved in a manner similar to that of reference 1.

TABLE VII. - 2^5 DESIGN EXPERIMENT ON PENICILLIN

Analysis	Prior guess of number of null mean squares, η	Prior type 1 risk, $\bar{R}_1(\eta)$	Assumed type 2 risk, $\bar{R}_2(\lambda)$	Number of mean squares pooled before test, m	Nominal size of preliminary pooling test, α_p	Nominal size of final significance test, α_f	Estimate of number of null mean squares, $\hat{\eta}$	Estimate of number of real effects, $\hat{\rho}$	Revised m	Estimate of average noncentrality parameter, $\hat{\lambda}$	Posterior type 1 risk, $\bar{R}_1(\eta)$	Posterior type 2 risk, $\bar{R}_2(\lambda)$
Significance point of view												
First	6	0.026	0.10	1	0.50	0.005	12	19	6	----	-----	------
Second	12	.105	.20	6	1.00	.001	20	11	---	21.5	[a]0.10	[a]0.04
Screening point of view												
First	6	0.08	0.10	1	0.50	0.01	6	25	3	----	-----	------
Second	6	.19	.05	3	1.00	.01	12	19	---	51.1	[b]0.18	[b]0.012

[a]Minor extrapolation.
[b]Major extrapolation.

As summarized in table VII, the analysis from a significance point of view with preferred m = 6 led to the conclusions that $\hat{\eta} = 20$ and $\hat{\rho} = 11$. The posterior risks were estimated at $\bar{R}_1(\eta) = 0.10$ and $\bar{R}_2(\lambda) = 0.04$. The analysis from a screening point of view led to the conclusions that $\hat{\eta} = 12$ and $\hat{\rho} = 19$. The posterior estimates of risk were off the curves, but extrapolated values are (very roughly) $\bar{R}_1(\eta) = 0.20$ and $\bar{R}_2(\lambda) = 0.01$. The only meaning that can be attached to such extrapolated risk estimates is that no more than the 19 largest mean squares should be regarded as being significant. The significance point of view with posterior $\bar{R}_1(\eta) = 0.10$ and $\bar{R}_2(\lambda) = 0.04$ gave $\hat{\rho} = 11$, which seems to be a reasonable appraisal of the experiment.

Davies' analysis (ref. 1, example 9.2, pp. 383 to 387 and pp. 416 to 418) is an attempt to test significance at the 5-percent level with the use of the three- and four-factor interactions (15 degrees of freedom) to estimate error variance. (The five-factor interaction was not pooled into the error estimate because it had been confounded with a block effect.) Davies concluded that the significant effects were μ_1, μ_3, μ_5, and μ_{35}; however, he felt that the conclusion with respect to μ_{35} was only weakly supported by the present experiment and was mainly supported by prior experiments. Based on the pooled three- and four-factor interactions, the error variance was estimated at $\hat{\sigma}^2 = 0.0034$. The upper 0.05 percent point of the F-distribution with 1 and 15 degrees of freedom is 4.54, whereas the 0.10 point is 3.07. Therefore at the 0.05 level, the mean squares that should be considered significant are all those larger than (0.0034)(4.54) = 0.01544, whereas at the 0.10 level, mean squares should be considered significant if they exceed (0.0034)(3.07) = 0.01044. At these levels (see table 9C.1 of ref. 1) the significant effects are as follows:

0.05 Level:	0.10 Level:
μ_5	μ_5
μ_1	μ_1
μ_3	μ_3
μ_{35}	μ_{35}
μ_{12345} (blocks)	μ_{12345} (blocks)
	μ_{12}
	μ_{135}
	μ_{1234}

The use of the F-test with 1 and 15 degrees of freedom is equivalent to the use of the double tailed t-test with 15 degrees of freedom. In essence, $\hat{\eta} = 23$ and $\hat{\rho} = 8$ are conclusions of the double tailed t-test with $\alpha = 0.10$. Usage of the double tailed t-test with $\alpha = 0.10$ is equivalent to testing the absolute values of the $\hat{\mu}_i$ using a single tailed t-test with $\alpha = 0.05$. For such a test, and with 15 degrees of freedom, tabulated data from reference 11 was used to plot the type 2 error curve, as shown by figure 19. This curve was then used with the estimated values of the eight largest $\hat{\mu}_i$ (aside from the grand mean) to compute a weighted average value of a posterior estimate of a type 2 risk in accordance with equations (3), (5), (6), and (14). The result was $\overline{R}_2(\lambda) = 0.060$.

The chain pooling ANOVA and the use of the t-test with Davies' estimate of σ^2 are then comparable at $\alpha = \overline{R}_1(\eta) = 0.10$. The comparison is as follows:

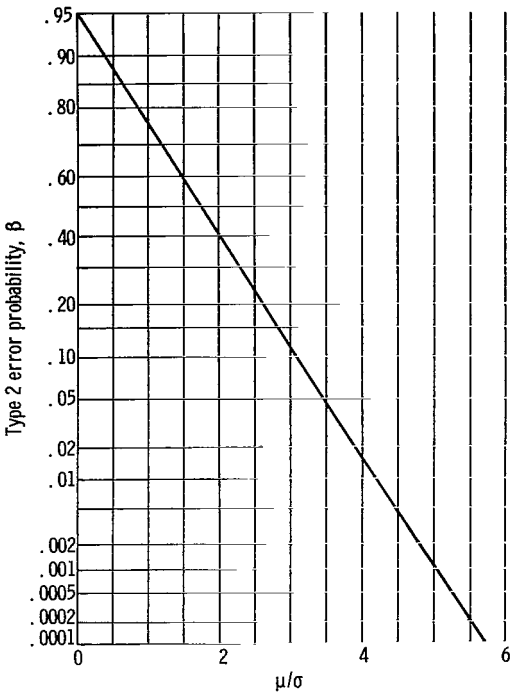


Figure 19. - Type 2 error probability for single tail t-test of size $\alpha = 0.05$ with 15 degrees of freedom (ref. 11).

Chain pooling	t-Test
$\overline{R}_1(\eta) = 0.10$	$\alpha = 0.10$
$m = 6$	d.f. = 15
$\hat{\eta} = 20$	$\hat{\eta} = 23$
$\hat{\rho} = 11$	$\rho = 8$
$\overline{R}_2(\lambda) = 0.04$	$\overline{R}_2(\lambda) = 0.060$

Thus with the same type 1 risk, the chain pooling has operated with a lower type 2 risk and declared three more effects significant, as compared with the t-test.

REFERENCES

1. Davies, Owen L., ed.: Design and Analysis of Industrial Experiments. Second ed., Hafner Publishing Co., 1956.

2. Daniel, Cuthbert: Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments. Technometrics, vol. 1, no. 4, Nov. 1959, pp. 311-341.

3. Wilk, M. B.; Gnanadesikan, R.; and Freeny, Anne E.: Estimation of Error Variance from Smallest Ordered Contrasts. J. Am. Stat. Assoc., vol. 58, no. 301, Mar. 1963, pp. 152-160.

4. Birnbaum, Allen: On the Analysis of Factorial Experiments Without Replication. Technometrics, vol. 1, no. 4, Nov. 1959, pp. 343-357.

5. Holms, Arthur G.: Designs of Experiments as Telescoping Sequences of Blocks for Optimum Seeking (As Intended for Alloy Development). NASA TN D-4100, Aug. 1967.

6. Kendall, Maurice G.; and Stuart, Alan: Inference and Relationship. Vol. 2 of The Advanced Theory of Statistics. Hafner Publishing Co., 1961.

7. Cochran, W. G.: The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total. Annals of Eugenics, vol. 11, 1941, pp. 47-52.

8. Eisenhart, Churchill; Hastay, Millard W.; and Wallis, W. Allen; eds.: Selected Techniques of Statistical Analysis for Scientific and Industrial Research, and Production and Management Engineering. McGraw-Hill Book Co., Inc., 1947, pp. 385-394.

9. Holms, Arthur G.: Multiple-Decision Procedures for the ANOVA of Two-Level Factorial Replication-Free Experiments. PhD Thesis, Western Reserve University, June, 1966. (Available from University Microfilms, Inc., Ann Arbor, Mich.)

10. Harter, Leon H.: Expected Values of Exponential, Weibull, and Gamma Order Statistics. Rep. No. ARL 64-31, Aerospace Research Labs., Wright-Patterson AFB, Feb. 1964. (Available from DDC as AD-436763).

11. Neyman, J.: Statistical Problems in Agricultural Experimentation. J. Roy. Stat. Soc., Suppl., vol. 2, pt. 2, 1935, pp. 107-180.

NASA-Langley, 1967 —— 19 E-3952

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546